

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2002, 18:26:03 ; Search time 30 seconds  
(without alignments)  
1995.208 Million cell updates/sec

Title: US-09-886-041-2

Perfect score: 346

Sequence: 1 MYNGSCRIEGTISQVMP.....ANSFQSDGQWDPHVH 346

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 562222 seqs, 172994929 residues

Word size : 0

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL\_19:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rvirus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	346	100.0	346	4 Q9BXC0	Q9bxc0 homo sapien
2	13	3.8	360	11 Q9EP66	Q9ep66 mus musculu
3	9	2.6	372	4 Q9HIC0	Q9hic0 homo sapien
4	9	2.6	391	5 O44148	O44148 caenorhabdi
5	8	2.3	121	8 Q9B9D7	Q9b9d7 drepana lac
6	8	2.3	227	6 Q9GLF9	Q9glf9 ovis aries
7	8	2.3	312	11 Q9IZC1	Q9izcl mus musculu
8	8	2.3	321	11 Q9IZB8	Q9izb8 mus musculu
9	8	2.3	396	2 Q9RH57	Q9rh57 bradyrhizob
10	8	2.3	572	3 Q9P810	Q9p810 emericeila
11	8	2.3	572	5 O76835	O76835 caenorhabdi
12	7	2.0	26	12 Q9ORV0	Q9orv0 hepatitis c
13	7	2.0	60	5 Q9G3F8	Q9g3f8 mesobuthus
14	7	2.0	60	5 Q9SP89	Q9sp89 mesobuthus
15	7	2.0	75	2 Q53417	Q53417 bradyrhizob
16	7	2.0	83	5 Q23304	Q23304 caenorhabdi

17	7	2.0	83	12 Q9Q3H4	Q9q3h4 hepatitis c
18	7	2.0	101	10 Q9AXA1	Q9axa1 oryza sativ
19	7	2.0	120	16 Q9HZ37	Q9hz37 pseudomonas
20	7	2.0	124	10 Q9ZTM2	Q9ztm2 petunia hyb
21	7	2.0	125	10 Q943Y7	Q943y7 thnopyrum
22	7	2.0	145	17 Q973Q2	Q973q2 sulfolobus
23	7	2.0	164	13 Q9DDV3	Q9ddv3 xenopus lae
24	7	2.0	166	10 Q940N3	Q940n3 arabidopsis
25	7	2.0	172	16 Q25061	Q25061 helicobacte
26	7	2.0	172	16 Q9ZME8	Q9zme8 helicobacte
27	7	2.0	174	2 Q93QI3	Q93qi3 corynebacte
28	7	2.0	186	2 Q9X3P7	Q9x3p7 caldicellul
29	7	2.0	204	16 Q9CJQ2	Q9cj2q2 pasteurella
30	7	2.0	206	16 Q9RTH2	Q9rth2 deinococcus
31	7	2.0	216	5 Q9W560	Q9w560 drosophila
32	7	2.0	242	4 Q60362	Q60362 homo sapien
33	7	2.0	242	16 Q9CKB8	Q9ckb8 pasteurella
34	7	2.0	249	2 Q93P96	Q93p96 microscilla
35	7	2.0	265	10 Q9M5K0	Q9m5k0 oryza sativ
36	7	2.0	272	5 Q9TZI3	Q9tzi3 caenorhabdi
37	7	2.0	275	5 Q9VPT7	Q9vpt7 drosophila
38	7	2.0	281	16 Q9PAJ8	Q9paj8 xylella fas
39	7	2.0	284	16 Q97JMI	Q97jmi clostridium
40	7	2.0	289	16 Q9I3I5	Q9i3i5 pseudomonas
41	7	2.0	291	13 Q9IAK3	Q9iak3 xenopus lae
42	7	2.0	300	12 Q9Q8F9	Q9q8f9 myxoma viru
43	7	2.0	306	10 Q9LV43	Q9lv43 arabidopsis
44	7	2.0	310	11 Q91ZB7	Q91zb7 mus musculu
45	7	2.0	314	2 Q9AKA9	Q9aka9 rickettsia

#### ALIGNMENTS

RESULT 1

Q9BXC0	PRELIMINARY;	PRT;	346 AA.
ID	Q9BXC0		
AC	Q9BXC0;		
DT	01-JUN-2001 (TrEMBLrel. 17, Created)		
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)		
DE	PUTATIVE CHEMOKINE RECEPTOR (G PROTEIN-COUPLED RECEPTOR).		
GN	FKSG80 OR GPR81.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Wang Y.-g, Gong L.;		
RT	"Molecular cloning of FKSG80, a novel gene encoding a putative chemokine receptor."		
RL	Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=21458557; PubMed=11574155;		
RA	Lee D.K., Nguyen T., Lynch K.R., Cheng R., Vanti W.B., Arkhitko O.,		
RA	Lewis T., Evans J.F., George S.R., O'Dowd B.F.;		
RT	"Discovery and mapping of ten novel G protein-coupled receptor genes."		
RL	Gene 275:83-91(2001).		
DR	EMBL; AF345568; AAK29071.1; -		
DR	EMBL; AF411110; AAL26481.1; -		
DR	HSSP; P34996; 1DDD.		
DR	InterPro; IPR000276; GPCR_Rhodpsn.		
DR	Pfam; PF00001; 7tm.1; 1.		
DR	PRINTS; PR00237; GPCRHHODPSN.		
DR	PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; UNKNOWN_1.		
DR	PROSITE; PS0262; G_PROTEIN_RECEP_F1_2; 1.		
KW	Receptor.		
SQ	SEQUENCE 346 AA; 39295 MW; E0DB114EEB3A47A5 CRC64;		
Query Match	100.0%;	Score 346;	DB 4; Length 346;

Best Local Similarity 100.0%; Pred. No. 0; Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYNSCCRIEGDITISQVMPPLLIIVAFVLGALNGVALCGFCFHKTKWKPSTVYLFLNLAVA 60  
 DB 1 MYNSCCRIEGDITISQVMPPLLIIVAFVLGALNGVALCGFCFHKTKWKPSTVYLFLNLAVA 60

QY 61 DFLMLICLPFTDYLLRRRHAFGDIPIRCVGLFTLAMNRAGSIIVLTVAADRYFKVVHP 120  
 DB 61 DFLMLICLPFTDYLLRRRHAFGDIPIRCVGLFTLAMNRAGSIIVLTVAADRYFKVVHP 120

QY 121 HHAVNTISTRVAAGICTLMAVLITGVYLLLENHLCVQETAVSCSFIMESANGWHDIM 180  
 DB 121 HHAVNTISTRVAAGICTLMAVLITGVYLLLENHLCVQETAVSCSFIMESANGWHDIM 180

QY 181 FOLEFFPLGLILCSFKIVMSLRRQOLARQARKKATREIMVAIVFTCYLPSVSAR 240  
 DB 181 FOLEFFPLGLILCSFKIVMSLRRQOLARQARKKATREIMVAIVFTCYLPSVSAR 240

QY 241 LYFLWTVPSSACDPSVHGALHITLSFTYMSMLDPLVYFSSPFPKYNKIKICSCLKPK 300  
 DB 241 LYFLWTVPSSACDPSVHGALHITLSFTYMSMLDPLVYFSSPFPKYNKIKICSCLKPK 300

QY 301 QPGHSKTQRPPEMPSINLGRSCISVANSFQSDGQWDPHIVEMH 346  
 DB 301 QPGHSKTQRPPEMPSINLGRSCISVANSFQSDGQWDPHIVEMH 346

RESULT 2

Q9EP66 PRELIMINARY; PRT; 360 AA.

AC Q9EP66;  
 DT 01-MAR-2001 (Tremblrel. 16, Created)  
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE PUTATIVE SEVEN TRANSMEMBRANE SPANNING RECEPTOR.  
 GN PUMAG OR PUMA-G.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=129/SVJ, AND C57BL/6;  
 RA Schaub A., Futterer A., Pfeiffer K.;  
 RT "PUMA-G, an interferon-gamma inducible gene in macrophages is a novel member of the seven transmembrane spanning superfamily.";  
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
 DR EMBL: AJ300199; CAC17791.1;  
 DR EMBL: AJ300198; CAC17790.1;  
 DR HSSP: P34996; IDDD.  
 DR MGD: MGI:1933383; Pumag.  
 DR InterPro: IPR000276; GPCR\_Rhodpsn.  
 DR Pfam: PF00001; 7tm.1; 1.  
 DR PRINTS: PR00237; GPCRHHODPSN  
 DR PROSITE: PS00237; G\_PROTEIN\_RECEP\_FL1; 1.  
 DR PROSITE: PS0262; G\_PROTEIN\_RECEP\_FL2; 1.  
 KW G-protein coupled receptor; Glycoprotein; Receptor; Transmembrane.  
 SQ SEQUENCE 360 AA; 41400 MW; CCCES2A247577FC CRC64;

Query Match 3.8%; Score 13; DB 11; Length 360;  
 Best Local Similarity 100.0%; Pred. No. 5.9e-05; Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 263 TLSFTYMSMLDP 275  
 DB 276 TLSFTYMSMLDP 288

RESULT 3  
 Q9H1C0

ID Q9H1C0 PRELIMINARY; PRT; 372 AA.  
 AC Q9H1C0;  
 DT 01-MAR-2001 (Tremblrel. 16, Created)  
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE PUTATIVE G PROTEIN-COUPLED RECEPTOR 92.  
 GN GRP92 OR GPR93.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=20517346; PubMed=11062477;  
 RA White K.E., Evans W.E., O'Riordan J.L.H., Spear M.C., Econs M.J.,  
 RA Lorenz-Depiereux B., Grabowski M., Meitinger T., Strom T.M.;  
 RT "Autosomal dominant hypophosphataemic rickets is associated with mutations in FGF23";  
 RL Nat. Genet. 26:345-348(2000).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=21458557; PubMed=11574155;  
 RA Lee D.K., Nguyen T., Lynch K.R., Cheng R., Vanti W.B., Arkhitko O.,  
 RA Lewis T., Evans J.F., George S.R., O'Dowd B.F.;  
 RT "Discovery and mapping of ten novel G protein-coupled receptor genes";  
 RL Gene 275:83-91(2001).  
 DR EMBL: AJ272207; CAC03715.1;  
 DR EMBL: AF411112; AAL26483.1;  
 DR HSSP: P34996; IDDD.  
 DR InterPro: IPR000276; GPCR\_Rhodpsn.  
 DR Pfam: PF00001; 7tm.1; 1.  
 DR PRINTS: PR00237; GPCRHHODPSN  
 DR PROSITE: PS00237; G\_PROTEIN\_RECEP\_FL1; UNKNOWN\_1.  
 DR PROSITE: PS0262; G\_PROTEIN\_RECEP\_FL2; 1.  
 KW Receptor.  
 SQ SEQUENCE 372 AA; 41346 MW; BA35709084BB6D84 CRC64;

Query Match 2.6%; Score 9; DB 4; Length 372;  
 Best Local Similarity 100.0%; Pred. No. 0.95; Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 273 LDPLVYVFS 281  
 DB 292 LDPLVYVFS 300

RESULT 4

O44148 PRELIMINARY; PRT; 391 AA.  
 ID O44148  
 AC O44148;  
 DT 01-JUN-1998 (Tremblrel. 06, Created)  
 DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE HYPOTHETICAL 45.5 KDA PROTEIN.  
 GN C49A9.7.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=BRISTOL N2;  
 RA MEDLINE=99069613; PubMed=9851916;  
 RA None;  
 RT "Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium.";  
 RL Science 282:2012-2018(1998).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=BRISTOL N2;  
 RA Fulton B., Wohlmann P.;  
 RT "The sequence of C. elegans cosmid C49A9.";

RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Waterston R.;  
 RT "Direct Submission.";  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF036693; AAK29785.1; -;  
 DR HSSP; P02699; 1F88.  
 DR InterPro; IPR000276; GPCR\_Rhodopsn.  
 DR Pfam; PF00001; 7tm\_1; 1.  
 DR PRINTS; PR00237; GPCRHOOPS.  
 DR PROSITE; PS00237; G\_PROTEIN\_RECEPTOR\_F1\_1; UNKNOWN\_1.  
 DR PROSITE; PS0262; G\_PROTEIN\_RECEPTOR\_F1\_2; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 391 AA; 45467 MW; D1448695E9AEBDE5 CRC64;

Query Match 2.3%; Score 9; DB 5; Length 391;  
 Best Local Similarity 100.0%; Pred. No. 0.99;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 YLFNLAVAD 61  
 |||||  
 Db 99 YLFNLAVAD 107

RESULT 5  
 Q9B9D7 PRELIMINARY; PRT; 121 AA.  
 AC Q9B9D7  
 DT 01-JUN-2001 (TReMBLrel. 17, Created)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)  
 DT 01-OCT-2001 (TReMBLrel. 18, Last annotation update)  
 DE NADH DEHYDROGENASE SUBUNIT 1 (FRAGMENT).  
 GN NDL.  
 OS Drepana lacertinaria.  
 OG Mitochondrion.  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;  
 OC Drepanoidea; Drepanidae; Drepaninae; Drepana.  
 OX NCBI\_TaxID=104429;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Abraham D., Ryrholm N., Wittzell H., Scoble M.J., Holloway J.D.,  
 RA Lofstedt C.;  
 RT "Molecular phylogeny of the subfamilies in Geometridae (Geometroidea: Lepidoptera).";  
 RL Mol. Phylogenet. Evol. 0:0-0(2001).  
 DR EMBL; AF178858; AAK00985.1; -;  
 DR InterPro; IPR001694; Resp\_chain\_NADH\_DH1.  
 DR Pfam; PF00146; NADHdh; 1.  
 DR PROSITE; PS00667; COMPLEX1\_NDL\_1; 1.  
 KW Mitochondrion.  
 FT NON\_TER 1 121  
 FT NON\_TER 121 121  
 SQ SEQUENCE 121 AA; 13810 MW; AA3FDB33066DB765 CRC64;

Query Match 2.3%; Score 8; DB 8; Length 121;  
 Best Local Similarity 100.0%; Pred. No. 3.9;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 322 SCISVANS 329  
 |||||  
 Db 112 SCISVANS 119

RESULT 6  
 Q9GLF9 PRELIMINARY; PRT; 227 AA.  
 AC Q9GLF9  
 DT 01-MAR-2001 (TReMBLrel. 16, Created)  
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)

DE C5A ANAPHYLATOXIN RECEPTOR (FRAGMENT).  
 OS Ovis aries (Sheep).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Caprinae; Ovis.  
 OX NCBI\_TaxID=9940;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Cain S.A., Woodruff T.M., Taylor S.M., Fairlie D.P., Sanderson S.D.,  
 RA Monk P.N.;  
 RT "Mutation of the first extracellular loop of the human C5a receptor identifies separate agonist and antagonist binding sites.";  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF284499; AAG12475.1; -;  
 DR InterPro; IPR000276; GPCR\_Rhodopsn.  
 DR Pfam; PF00001; 7tm\_1; 1.  
 DR PRINTS; PR00237; GPCRHOOPS.  
 DR PROSITE; PS00237; G\_PROTEIN\_RECEPTOR\_F1\_1; UNKNOWN\_1.  
 DR PROSITE; PS0262; G\_PROTEIN\_RECEPTOR\_F1\_2; 1.  
 KW Receptor.  
 FT NON\_TER 1 227  
 FT NON\_TER 227 227  
 SQ SEQUENCE 227 AA; 25443 MW; F6D85AB29FABF6A1 CRC64;

Query Match 2.3%; Score 8; DB 6; Length 227;  
 Best Local Similarity 100.0%; Pred. No. 6.8;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 NLAVADFL 63  
 |||||  
 Db 15 NLAVADFL 22

RESULT 7  
 Q91ZC1 PRELIMINARY; PRT; 312 AA.  
 AC Q91ZC1  
 DT 01-DEC-2001 (TReMBLrel. 19, Created)  
 DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
 DE G. PROTEIN-COUPLED RECEPTOR.  
 GN MRG3.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6;  
 RX MEDLINE=21435808; PubMed=11551509;  
 RA Dong X., Han S.-K., Zylka M.J., Simon M.I., Anderson D.J.;  
 RT "A Diverse Family of GPCRs Expressed in Specific Subsets of Nociceptive Somatosensory Neurons.";  
 RL Cell 106:619-632(2001).  
 DR EMBL; AY042201; AAK91797.1; -;  
 KW Receptor.  
 SQ SEQUENCE 312 AA; 35509 MW; D71FF8F6008B2129 CRC64;

Query Match 2.3%; Score 8; DB 11; Length 312;  
 Best Local Similarity 100.0%; Pred. No. 9.1;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 NLAVADFL 63  
 |||||  
 Db 67 NLAVADFL 74

RESULT 8  
 Q91ZB8 PRELIMINARY; PRT; 321 AA.  
 ID Q91ZB8  
 AC Q91ZB8;  
 DT 01-DEC-2001 (TReMBLrel. 19, Created)  
 DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)

```

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE G PROTEIN-COUPLED RECEPTOR.
GN MRGD.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RX MEDLINE=21435808; PubMed=11551509;
RA Dong X., Han S.-K., Zylka M.J., Simon M.I., Anderson D.J.;
RT "A Diverse Family of GPCRs Expressed in Specific Subsets of
RL Nociceptive Somatosensory Neurons.";
Cell 106:619-632(2001).
DR EMBL; AY042209; AAK91800.1; -.
KW Receptor.
SQ SEQUENCE 321 AA; 36125 MW; B0642547A75077B3 CRC64;

Query Match 2.3%; Score 8; DB 11; Length 321;
Best Local Similarity 100.0%; Pred. No. 9.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 56 NLAVADFL 63
Db 67 NLAVADFL 74

RESULT 9
Q9RH57 PRELIMINARY; PRT; 396 AA.
AC Q9RH57;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE ADENYLATE CYCLASES-LIKE PROTEIN.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Bradyrhizobium group; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110SPC4;
RA Mueller P., Stingel D.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF065159; AAF22882.1; -.
SQ SEQUENCE 396 AA; 43188 MW; CB729E48EED6DC27 CRC64;

Query Match 2.3%; Score 8; DB 2; Length 396;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 203 LRRRQOLA 210
Db 24 LRRRQOLA 31

RESULT 10
Q9P810 PRELIMINARY; PRT; 572 AA.
AC Q9P810;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE DELTA-1-PYRROLINE-5-CARBOXYLATE DEHYDROGENASE.
GN PRNC.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=5072;
RN [1]
RP SEQUENCE FROM N.A.
RA Demais S., Gavrias V., Scazzocchio C.;

```

```

RT "Primary structure of the nuclear prnC gene involved in the
RT mitochondrial pathway for proline utilization in Aspergillus
RT nidulans.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF252630; AAF72527.1; -.
DR InterPro: IPR002086; Aldehyde_dehydr.
DR Pfam; PF00171; algedh; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; UNKNOWN_1.
SQ SEQUENCE 572 AA; 61921 MW; 1EB410940931C71E CRC64;

Query Match 2.3%; Score 8; DB 3; Length 572;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 127 ISTRVAAG 134
Db 302 ISTRVAAG 309

RESULT 11
O76835 PRELIMINARY; PRT; 572 AA.
ID O76835
AC O76835;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 65.2 KDA PROTEIN.
GN T12A2.15.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Latreille P.;
RT "The sequence of C. elegans cosmid T12A2.";
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U13019; AAC24452.1; -.
DR HSSP; P04410; 1A25.
DR InterPro; IPR000008; C2.
DR Pfam; PF00168; C2; 2.
DR PRINTS; PR00360; C2DOMAIN.
DR SMART; SM00239; C2; 2.
DR PROSITE; PS50004; C2_DOMAIN_2; 2.
KW Hypothetical protein.
SQ SEQUENCE 572 AA; 65171 MW; A5B288895BD5B09D CRC64;

Query Match 2.3%; Score 8; DB 5; Length 572;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 84 GDIPCRVG 91
Db 165 GDIPCRVG 172

RESULT 12

```



Q9QRV0 Q9QRV0 PRELIMINARY; PRT; 26 AA.  
AC Q9QRV0;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)  
DE E2 GLYCOPROTEIN HYPERVARIABLE REGION (FRAGMENT).  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=11103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A-AS;  
RA Yeh C.-T.;  
RT "Replication of hepatitis C virus in the ascitic mononuclear cells and development of distinct quasiespecies in the ascitic fluid."  
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF109739; AAD51570.1; -  
FT NON\_TER 1 1  
FT 26 26  
SQ SEQUENCE 26 AA; 2462 MW; 945C5E047695C5CC CRC64;  
  
Query Match 2.0%; Score 7; DB 12; Length 26;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 90 VGLFTLA 96  
Db 17 VGLFTLA 23  
  
RESULT 13  
Q963F8 Q963F8 PRELIMINARY; PRT; 60 AA.  
AC Q963F8;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE TOXIN TKXS1.  
OS Mesobuthus martensii (Manchurian scorpion) (Buthus martensii).  
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Scorpiones;  
OC Buthoidea; Buthidae; Buthus.  
OX NCBI\_TaxID=34649;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Shunyi Z., Wenxin L.;  
RT "Genomic DNA encoding BmTKXS1, a new K+ channel toxin derived from the scorpion Buthus martensii Karsch, containing an intron of 75 base pairs."  
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF380940; AAK58091.1; -  
SQ SEQUENCE 60 AA; 6523 MW; 6F6F5F57328B0BC5 CRC64;  
  
Query Match 2.0%; Score 7; DB 5; Length 60;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 131 VAAGIVC 137  
Db 26 VAAGIVC 32  
  
RESULT 14  
Q95P89 Q95P89 PRELIMINARY; PRT; 60 AA.  
AC Q95P89;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE PUTATIVE POTASSIUM ION CHANNEL BLOCKER TKXS1.  
OS Mesobuthus martensii (Manchurian scorpion) (Buthus martensii).  
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Scorpiones;

OC Buthoidea; Buthidae; Buthus.  
OX NCBI\_TaxID=34649;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=VENOM GLAND;  
RA Zhu S., Li W.;  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF155369; AAK61819.1; -  
SQ SEQUENCE 60 AA; 6523 MW; 7AB56F57328B11EC CRC64;  
  
Query Match 2.0%; Score 7; DB 5; Length 60;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 131 VAAGIVC 137  
Db 26 VAAGIVC 32  
  
RESULT 15  
Q53417 Q53417 PRELIMINARY; PRT; 75 AA.  
AC Q53417;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)  
DE NOLZ.  
GN NOLZ.  
OS Bradyrhizobium japonicum.  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Bradyrhizobium group; Bradyrhizobium.  
OX NCBI\_TaxID=375;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94281668; PubMed=8012039;  
RA Dockendorff T.C., Sharma A.J., Stacey G.;  
RT "Identification and characterization of the nolyz genes of Bradyrhizobium japonicum."  
RL Mol. Plant Microbe Interact. 7:173-180(1994).  
DR EMBL: S70736; AAB31118.2; -  
SQ SEQUENCE 75 AA; 8662 MW; 666068F4C2684CDA CRC64;  
  
Query Match 2.0%; Score 7; DB 2; Length 75;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 235 PSVSARL 241  
Db 4 PSVSARL 10  
  
RESULT 16  
Q23304 Q23304 PRELIMINARY; PRT; 83 AA.  
AC Q23304;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)  
DE ZC412.8 PROTEIN.  
GN H12D21.3.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Ainscough R.;  
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA McMurray A.;  
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL: Z78067; CAB01527.1; -

DR EMBL; 292849; CAB07424.1; -.  
SQ SEQUENCE 83 AA; 8004 MW; 8C8869460996384F CRC64;

Query Match 2.0%; Score 7; DB 5; Length 83;  
Best Local Similarity 100.0%; Pred. No. 31;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 LLIVAFV 27  
| | | | | | |  
DB 10 LLIVAFV 16

## RESULT 17

Q9Q3H4 ID Q9Q3H4 PRELIMINARY; PRT; 83 AA.  
AC Q9Q3H4;  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE GENOME POLYPROTEIN [CONTAINS: ENVELOPE GLYCOPROTEIN E1 (GP35)]  
DE (FRAGMENT).  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=11103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=30A;  
RX MEDLINE=21373979; PubMed=11481629;  
RA Lin H.J., Seeff L.B., Barbosa L., Hollinger F.B.;  
RT "Occurrence of identical hypervariable region 1 sequences of hepatitis C virus in transfusion recipients and their respective blood donors: Divergence over time.";  
RL Hepatology 34:424-429(2001).  
CC -1- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.  
DR EMBL; AF206449; AAF19961.1; -.  
DR InterPro; IPR002519; HCV\_env.  
DR Pfam; PF01539; HCV\_env; 1.  
KW Coat protein; Envelope protein; Glycoprotein; Polyprotein;  
KW Transmembrane.  
FT NON\_TER 1 1  
FT NON\_TER 83 83  
SQ SEQUENCE 83 AA; 8711 MW; FF1C75E8FCFBBA5 CRC64;

Query Match 2.0%; Score 7; DB 12; Length 83;  
Best Local Similarity 100.0%; Pred. No. 31;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 90 VGLFTLA 96  
| | | | | | |  
DB 72 VGLFTLA 78

## RESULT 18

Q9AXA1 ID Q9AXA1 PRELIMINARY; PRT; 101 AA.  
AC Q9AXA1;  
DT 01-JUN-2001 (Tremblrel. 17, Created)  
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
DE P0501G01.16 PROTEIN.  
GN P0501G01.16.  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzoae; Oryza.  
OX NCBI\_TaxID=4530;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV\_NIPPONBARE;  
RA Sasaki T., Matsumoto T., Yamamoto K.;  
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC clone:P0501G01.";

RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AP002819; BAB21087.1; -.  
SQ SEQUENCE 101 AA; 10785 MW; E1575443B63597FE CRC64;

Query Match 2.0%; Score 7; DB 10; Length 101;  
Best Local Similarity 100.0%; Pred. No. 37;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 204 RRRQOLA 210  
| | | | | | |  
DB 39 RRRQOLA 45

## RESULT 19

Q9HZ37 ID Q9HZ37 PRELIMINARY; PRT; 120 AA.  
AC Q9HZ37;  
DT 01-MAR-2001 (Tremblrel. 16, Created)  
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
DT 01-OCT-2001 (Tremblrel. 18, Last annotation update)  
DE HYPOTHETICAL PROTEIN PA3203.  
GN PA3203.  
OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
OX NCBI\_TaxID=287;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 15692 / PAOI;  
RX MEDLINE=20437337; PubMed=10984043;  
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,  
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
RA Garber R.L., Goltz L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;  
RT "Complete genome sequence of Pseudomonas aeruginosa PAOI, an opportunistic pathogen.";  
RL Nature 406:959-964(2000).  
DR EMBL; AE004744; AAG06591.1; -.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 120 AA; 13479 MW; 451B7F7C2A6DF475 CRC64;

Query Match 2.0%; Score 7; DB 16; Length 120;  
Best Local Similarity 100.0%; Pred. No. 43;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 AFVLGAL 31  
| | | | | | |  
DB 101 AFVLGAL 107

## RESULT 20

Q9ZTM2 ID Q9ZTM2 PRELIMINARY; PRT; 124 AA.  
AC Q9ZTM2;  
DT 01-MAY-1999 (Tremblrel. 10, Created)  
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE PGPS/NH20 (FRAGMENT).  
GN PGPS/NH20.  
OS Petunia hybrida (Petunia).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; euasterids I; Solanales; Solanaceae; Petunia.  
OX NCBI\_TaxID=4102;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=GERMINATING PETUNIA POLLEN TREATED WITH KAEMPFEROL;  
RX MEDLINE=20317212; PubMed=10859200;  
RA Guyon V.N., Astwood J.D., Garner E.C., Dunker A.K., Taylor L.P.;  
RT "Isolation and characterization of cDNAs expressed in the early stages of flavonol-induced pollen germination in petunia.";

RL Plant Physiol. 123:699-710(2000).  
 DR EMBL; AF049936; AAD02560.1; -.  
 DR HSP; P25816; ICOA.  
 DR InterPro; IPR002097; Profilin.  
 DR Pfam; PF00235; profilin; 1.  
 DR PRINTS; PR00392; PROFILIN.  
 DR SMART; SM00392; PROF; 1.  
 FT NON\_TER 1  
 SQ SEQUENCE 124 AA; 13283 MW; 7716090AB96F8F1E CRC64;

Query Match 2.08; Score 7; DB 10; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 44;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 282 SPSPFKF 288  
 Db 29 SPSPFKF 35  
 |||||

## RESULT 21

Q943V7 PRELIMINARY; PRT; 125 AA.  
 AC Q943V7;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE INORGANIC PHOSPHATE TRANSPORTER (FRAGMENT).  
 GN PT.  
 OS Thinopyrum Intermedium.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;  
 OC Triticeae; Thinopyrum.  
 OX NCBI\_TaxID=85679;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RA Davies T.G.E., Ying J., Xu Q., Li Z., Gordon-Weeks R.;  
 RT "Analysis of high-affinity phosphate transporter expression in alien  
 RL translocation lines of Chinese winter wheats."  
 DR EMBL; AJ13964; CAC88705.1; -.  
 FT NON\_TER 125 125  
 SQ SEQUENCE 125 AA; 13348 MW; BCF6E0C47CB27122 CRC64;

Query Match 2.08; Score 7; DB 10; Length 125;  
 Best Local Similarity 100.0%; Pred. No. 45;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 33 NGVALCG 39  
 Db 37 NGVALCG 43  
 |||||

## RESULT 22

ID Q973Q2 PRELIMINARY; PRT; 145 AA.  
 AC Q973Q2;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE HYPOTHETICAL PROTEIN ST0845.  
 GN ST0845.  
 OS Sulfolobus tokodaii.

OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.  
 OX NCBI\_TaxID=111955;  
 RN [1]  
 RP STRAIN-JCM 10545 / 7;  
 RC PubMed=11572479;  
 RX Kawarabayasi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,  
 RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,  
 RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,  
 RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,

RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,  
 RA Oshima T., Kikuchi H.;  
 RT "Complete genome sequence of an aerobic thermophilic  
 RT Crenarchaeon, Sulfolobus tokodaii strain7."  
 RL DNA Res. 8:123-140(2001).  
 DR EMBL; AP000983; BAB65858.1; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 145 AA; 16705 MW; 962411A47EEAF8E1 CRC64;

Query Match 2.08; Score 7; DB 17; Length 145;  
 Best Local Similarity 100.0%; Pred. No. 51;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 102 SIVFLTV 108  
 Db 8 SIVFLTV 14  
 |||||

## RESULT 23

Q9DDV3 PRELIMINARY; PRT; 164 AA.  
 AC Q9DDV3;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE HOMEBOX PROTEIN SIX3.2 (FRAGMENT).  
 GN SIX3.2.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
 OC Xenopodinae; Xenopus.  
 OX NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RX MEDLINE=21152917; PubMed=11231090;  
 RA Ghanbari H., Seo H.C., Fjose A., Brandli A.W.;  
 RT "Molecular cloning and embryonic expression of Xenopus Six homeobox  
 RT genes."  
 RL Mech. Dev. 101:271-277(2001).  
 DR EMBL; AF276992; AAG42358.1; -.  
 DR InterPro; IPR001356; Homeobox.  
 DR InterPro; IPR000047; HTH\_repressr.  
 DR Pfam; PF00046; homeobox; 1.  
 DR PRINTS; PR00031; HTHREPRESSR.  
 DR SMART; SM00389; HOX; 1.  
 DR PROSITE; PS50071; HOMEBOX\_2; 1.  
 FT NON\_TER 164 164  
 SQ SEQUENCE 164 AA; 19270 MW; A41A86FF4ADB2D24 CRC64;

Query Match 2.08; Score 7; DB 13; Length 164;  
 Best Local Similarity 100.0%; Pred. No. 57;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 149 YLLLENH 155  
 Db 55 YLLLENH 61  
 |||||

## RESULT 24

Q940N3 PRELIMINARY; PRT; 166 AA.  
 AC Q940N3;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE AT4G17240/DLA655C.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsiis.  
 OX NCBI\_TaxID=3702;  
 RN [1]

```
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
RA Bowser L., Carncinci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jlang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis cDNA clones.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY054245; AAL06904.1; -.
SQ SEQUENCE 166 AA; 18304 MW; 8E1DECC5FAFE79B8 CRC64;

Query Match 2.0%; Score 7; DB 10; Length 166;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 191 IILFCSF 197
DB 42 IILFCSF 48

RESULT 25
O25061 PRELIMINARY; PRT; 172 AA.
AC O25061;
DT 01-JAN-1998 (TRENBLrel. 05, Created)
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE HYPOTHETICAL 20.1 KDA PROTEIN.
GN HP0287.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
pylori.";
RL Nature 388:539-547(1997).
DR EMBL; AE000547; AAD07359.1; -.
DR TIGR; HP0287; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 172 AA; 20110 MW; B0C8F44AA739566C CRC64;

Query Match 2.0%; Score 7; DB 16; Length 172;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 LEFFMPL 189
DB 150 LEFFMPL 156

RESULT 26
O25ME8 PRELIMINARY; PRT; 172 AA.
AC O25ME8;
DT 01-MAY-1999 (TRENBLrel. 10, Created)
DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
DT 01-MAY-2000 (TRENBLrel. 13, Last annotation update)
OX NCBI_TaxID=80339;
```

```
DE PUTATIVE.
GN JHP0272.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Gullid B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
gastric pathogen Helicobacter pylori.";
RL Nature 397:176-180(1999).
DR EMBL; AE001464; AAD05853.1; -.
KW Complete proteome.
SQ SEQUENCE 172 AA; 20296 MW; 8F340C659A1B4DF2 CRC64;

Query Match 2.0%; Score 7; DB 16; Length 172;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 LEFFMPL 189
DB 150 LEFFMPL 156

RESULT 27
O930I3 PRELIMINARY; PRT; 174 AA.
AC O930I3;
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE HYPOTHETICAL 17.0 KDA PROTEIN.
OS Corynebacterium equi (Rhodococcus equi).
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.
OX NCBI_TaxID=43767;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC6939;
RA Vanniasinkam T., Barton M.D., Heuzenroeder M.W.;
RT "Characterization of the groEL gene of Rhodococcus equi.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF233387; AAK95492.1; -.
KW Hypothetical protein.
SQ SEQUENCE 174 AA; 16967 MW; E0783D3628EF24E6 CRC64;

Query Match 2.0%; Score 7; DB 2; Length 174;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 315 ISNLGRR 321
DB 10 ISNLGRR 16

RESULT 28
O9X3P7 PRELIMINARY; PRT; 186 AA.
AC O9X3P7;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE HYPOTHETICAL 21.3 KDA PROTEIN (FRAGMENT).
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Thermoanaerobacter group; Caldicellulosiruptor.
OX NCBI_TaxID=80339;
```

```

RN  SEQUENCE FROM N.A.
RP  STRAIN=TOK7B.1;
RX  MEDLINE=201711169; PubMed=10706665;
RA  Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA  Bergquist P.L.;
RT  "Multidomain and multifunctional glycosyl hydrolases from the extreme
RL  thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL  Curr. Microbiol. 40:333-340(2000).
DR  EMBL: AF078737; AAD30366.1;
KW  Hypothetical protein.
FT  NON_TER 186 186
SQ  SEQUENCE 186 AA; 21323 MW; E8423BD1D0D02183 CRC64;

Query Match      2.0%; Score 7; DB 2; Length 186;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  224 VVAIVFI 230
Db  162 VVAIVFI 168

RESULT 29
Q9CJQ2
ID  Q9CJQ2 PRELIMINARY; PRT; 204 AA.
AC  Q9CJQ2;
DT  01-JUN-2001 (TrEMBLrel. 17, Created)
DT  01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT  01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE  HYPOTHETICAL PROTEIN PM1942.
GN  PM1942.
OS  Pasteurella multocida.
OC  Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC  Pasteurella.
OX  NCBI_TaxID=747;
RN  SEQUENCE FROM N.A.
RP  STRAIN=PM70;
RX  MEDLINE=21145866; PubMed=11248100;
RA  May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT  "Complete genomic sequence of Pasteurella multocida pm70.";
RL  Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR  EMBL: AE006231; AAK04026.1;
DR  IPrPro: IPR002524; Cation_efflux.
DR  Pfam: PF01545; Cation_efflux.2.
KW  Hypothetical protein; Complete proteome.
SQ  SEQUENCE 204 AA; 22451 MW; A96A5772469588FF CRC64;

Query Match      2.0%; Score 7; DB 16; Length 204;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  22 LIVAFVL 28
Db  176 LIVAFVL 182

RESULT 30
Q9RTH2
ID  Q9RTH2 PRELIMINARY; PRT; 206 AA.
AC  Q9RTH2;
DT  01-MAY-2000 (TrEMBLrel. 13, Created)
DT  01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT  01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE  HYPOTHETICAL 22.1 KDA PROTEIN.
GN  DRI1792.
OS  Deinococcus radiodurans.
OC  Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX  NCBI_TaxID=1299;
RN  SEQUENCE FROM N.A.
RP  STRAIN=R1;

```

```

RX  MEDLINE=20036896; PubMed=10567266;
RA  White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA  Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA  Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA  Vamatheva J.J., Lam P., McDonald L., Utterback T., Zaleski C.,
RA  Makareva K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA  Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA  Fraser C.M.;
RT  "Genome sequence of the radioresistant bacterium Deinococcus
RT  radiodurans R1.";
RL  Science 286:1571-1577(1999).
DR  EMBL: AE002020; AAF11349.1;
DR  TIGR; DRI792;
KW  Hypothetical protein; Complete proteome.
SQ  SEQUENCE 206 AA; 22146 MW; D8ABD279C1D8561C CRC64;

Query Match      2.0%; Score 7; DB 16; Length 206;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  104 VFLTVVA 110
Db  163 VFLTVVA 169

RESULT 31
Q9W560
ID  Q9W560 PRELIMINARY; PRT; 216 AA.
AC  Q9W560;
DT  01-MAY-2000 (TrEMBLrel. 13, Created)
DT  01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT  01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE  CG14811 PROTEIN.
GN  CG14811.
OS  Drosophila melanogaster (Fruit fly).
OC  Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC  Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC  Ephydroidea; Drosophilidae; Drosophila.
OX  NCBI_TaxID=7227;
RN  SEQUENCE FROM N.A.
RP  STRAIN=BERKELEY;
RX  MEDLINE=20196006; PubMed=107311132;
RA  Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA  Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA  George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA  Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA  Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA  Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA  Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA  Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA  Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA  Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA  Burtis K.C., Busam D.A., Butler H., Cadieu E., Canter A., Chandra I.,
RA  Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA  de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA  Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA  Durbin K.J., Evangelista C.C., Ferraz C., Ferrier A.S., Fleischmann W.,
RA  Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA  Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA  Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA  Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA  Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA  Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA  Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA  Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA  Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA  Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA  Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA  Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA  Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA  Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

```

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.:  
RT "The genome sequence of *Drosophila melanogaster*."  
DR EMBL: AE003422; AAF45667.1; -.  
DR FlyBase: FBgn0029590; CG14811.  
SQ SEQUENCE 216 AA; 24391 MW; A1B9DB0A86C0563A CRC64;

Query Match 2.0%; Score 7; DB 5; Length 216;  
Best Local Similarity 100.0%; Pred. No. 73;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 296 SLKPKQP 302  
Db 67 SLKPKQP 73  
|||||

RESULT 32  
O60362 PRELIMINARY; PRT; 242 AA.  
AC O60362;  
DT 01-AUG-1998 (TREMBlrel. 07, Created)  
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE HYPOTHETICAL 27.3 KDA PROTEIN.  
GN 44W2.1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99425270; PubMed=10493829;  
RA Loftus B.J., Kim U.J., Sneddon V.P., Kalush F., Brandon R.,  
RA Fuhrmann J., Mason T., Crosby M.L., Barnstead M., Cronin L.,  
RA Deslattes Mays A., Cao Y., Xu R.X., Kang H.L., Mitchell S.,  
RA Eichler E.E., Harris P.C., Venter J.C., Adams M.D.:  
RT "Genome duplications and other features in 12 Mb of DNA sequence from  
human chromosome 16p and 16q."  
RL Genomics 60:295-308(1999).  
DR EMBL: AC004381; AAC31666.1; -.  
DR InterPro: IPR004114; THUMP.  
DR Pfam: PF02926; THUMP; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 242 AA; 27310 MW; A0099D58349AF315 CRC64;

Query Match 2.0%; Score 7; DB 4; Length 242;  
Best Local Similarity 100.0%; Pred. No. 81;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 AGIVCTL 139  
Db 107 AGIVCTL 113  
|||||

RESULT 33  
Q9CKB8 PRELIMINARY; PRT; 242 AA.  
AC Q9CKB8;  
DT 01-JUN-2001 (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE DSB.  
GN DSB2 OR PM1706.  
OS Pasteurella multocida.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Pasteurella.  
OX NCBI\_TaxID=747;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=PM70;  
RX MEDLINE=21145866; PubMed=11248100;  
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.:  
RT "Complete genomic sequence of *Pasteurella multocida* Fm70."  
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).  
DR EMBL: AE006207; AAK03790.1; -.  
DR InterPro: IPR003834; Dsbd\_Dipz.  
DR Pfam: PF02683; Dsbd; 1.  
KW Complete proteome.  
SQ SEQUENCE 242 AA; 25653 MW; FD2EEBA66E49A36D CRC64;

Query Match 2.0%; Score 7; DB 16; Length 242;  
Best Local Similarity 100.0%; Pred. No. 81;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 90 VGLFTLA 96  
Db 41 VGLFTLA 47  
|||||

RESULT 34  
Q93P96 PRELIMINARY; PRT; 249 AA.  
AC Q93P96;  
DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE MS135, PUTATIVE ARYL SULFATASE.  
OS Microscilla sp. PREL.  
OG Plasmid pSD15.  
OC Bacteria; CFB group; Flexibacter group; Microscilla.  
OX NCBI\_TaxID=155537;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PREL;  
RA Zhong Z., Toukadarian A., Helinski D., Knauf V., Sykes S.,  
RA Wilkinson J.E., O'Bryne C., Shea T., De Loughery C., Caspi R.:  
RT "Sequence Analysis of a 101 kb Plasmid Required for Agar-degradation  
of a *Microscilla* strain."  
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF339846; AAK62857.1; -.  
KW Plasmid.  
SQ SEQUENCE 249 AA; 27900 MW; FCE354E91E948899 CRC64;

Query Match 2.0%; Score 7; DB 2; Length 249;  
Best Local Similarity 100.0%; Pred. No. 83;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 TLAMNRA 100  
Db 36 TLAMNRA 42  
|||||

RESULT 35  
Q9M5K0 PRELIMINARY; PRT; 265 AA.  
AC Q9M5K0;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
DE PHOSPHATE TRANSPORTER (FRAGMENT).  
OS *Oryza sativa* (Rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzeae; *Oryza*.  
OX NCBI\_TaxID=4530;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. INDICA;  
RA Yu F., Zhang A., Zhang F., Chen S.:  
RT "Rice phosphate transporter."  
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.

```

DR EMBL; AF229169; AAF40188.1; -.
FT NON_TER 1
FT NON_TER 265
SQ SEQUENCE 265 AA; 28329 MW; DC88DD1D52E06A74 CRC64;

Query Match
Best Local Similarity 2.0%; Score 7; DB 10; Length 265;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 NGVALCG 39
Db 43 NGVALCG 49
|||||||

RESULT 36
Q9T2I3 PRELIMINARY; PRT; 272 AA.
ID Q9T2I3
AC Q9T2I3
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE F59H5.3 PROTEIN.
GN F59H5.3
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lighning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkneen R.,
RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wooldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Jones K., Graves T., Antoniou B.;
RT "The sequence of C. elegans cosmid F59H5.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF098991; AAC67450.1; -.
DR InterPro; IPR002083; MATH.
DR Pfam; PF00917; MATH; 1.
DR SMART; SM00061; MATH; 1.
SQ SEQUENCE 272 AA; 31260 MW; 89421E02355F7241 CRC64;

Query Match
Best Local Similarity 2.0%; Score 7; DB 5; Length 272;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 189 LGIILFC 195
Db 49 LGIILFC 55
|||||||

RESULT 37
Q9VPT7 PRELIMINARY; PRT; 275 AA.
ID Q9VPT7
AC Q9PAJ8
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

```

```

AC Q9VPT7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE CG3876 PROTEIN.
GN CG3876.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abiril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foslter C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003588; AAF51454.1; -.
DR FlyBase; FBgn0031284; CG3876.
SQ SEQUENCE 275 AA; 31895 MW; CEAF8F33CFC340141 CRC64;

Query Match
Best Local Similarity 2.0%; Score 7; DB 5; Length 275;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 YLPSVSA 239
Db 60 YLPSVSA 66
|||||||

RESULT 38
Q9PAJ8 PRELIMINARY; PRT; 281 AA.
ID Q9PAJ8
AC Q9PAJ8
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

```

```

DE  HYPOTHETICAL PROTEIN XF2517.
CN  Xylella fastidiosa.
OS  Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC  Xylella.
OX  NCBI_TaxID=2371;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=9A5C;
RX  MEDLINE=20365717; PubMed=10910347;
RA  Simpson A.J.G.; Reinach F.C.; Arruda P.; Abreu F.A.; Acencio M.;
RA  Alvaranga R.; Alves L.M.C.; Araya J.E.; Baia G.S.; Baptista C.S.;
RA  Barros M.H.; Bonaccorsi E.D.; Bordin S.; Bove J.M.; Briones M.R.S.;
RA  Bueno M.R.P.; Camargo A.A.; Camargo L.E.A.; Carraro D.M.; Carrier H.;
RA  Colauto N.B.; Colombo C.; Costa F.F.; Costa M.C.R.; Costa-Neto C.M.;
RA  Coutinho L.L.; Cristofani M.; Dias-Neto E.; Docena C.; El-Dorry H.;
RA  Facincani A.P.; Ferreira J.S.; Ferreira V.C.A.; Ferro J.A.;
RA  Fraga J.S.; Franca S.C.; Franco M.C.; Frohme M.; Furian L.R.;
RA  Garnier M.; Goldman G.H.; Goldman M.H.S.; Gomes S.L.; Gruber A.;
RA  Ho P.L.; Hoheisel J.D.; Junqueira M.L.; Kemper E.L.; Kitajima J.P.;
RA  Krieger J.E.; Kuramae E.E.; Laigret F.; Lambais M.R.; Leite L.C.C.;
RA  Lemos E.G.M.; Lemos M.V.F.; Lopes S.A.; Lopes C.R.; Machado J.A.;
RA  Machado M.A.; Madeira A.M.B.N.; Madeira H.M.F.; Marino C.L.;
RA  Marques M.V.; Martins E.A.L.; Martins E.M.F.; Matsukuma A.Y.;
RA  Menck C.F.M.; Miracca E.C.; Miyaki C.Y.; Monteiro-Vitorello C.B.;
RA  Moon D.H.; Nagai M.A.; Nascimento A.L.T.O.; Netto L.E.S.;
RA  Nhani A. Jr.; Nobrega F.G.; Nunes L.R.; Oliveira M.A.;
RA  de Oliveira M.C.; de Oliveira R.C.; Palmieri D.A.; Paris A.;
RA  Peixoto B.R.; Pereira G.A.G.; Pereira H.A. Jr.; Pesquero J.B.;
RA  Quaggio R.B.; Roberto P.G.; Rodrigues V.; de Rosa A.J.M.;
RA  de Rosa V.E. Jr.; de Sa R.G.; Santelli R.V.; Sawasaki H.E.;
RA  da Silva A.C.R.; da Silva A.M.; da Silva F.R.; Silva W.A. Jr.;
RA  da Silveira J.F.; Silvestri M.L.Z.; Siqueira W.J.; de Souza A.A.;
RA  de Souza A.P.; Terenzi M.F.; Truffi D.; Tsai S.M.; Tsuchiko M.H.;
RA  Vallada H.; Van Sluys M.A.; Verjovski-Almeida S.; Vettore A.L.;
RA  Zago M.A.; Zatz M.; Meidanis J.; Setubal J.C.;
RT  "The genome sequence of the plant pathogen Xylella fastidiosa."
RL  Nature 406:151-159(2000).
DR  EMBL; AE004059; AAF85315.1; -.
KW  Hypothetical protein; Complete proteome.
SQ  SEQUENCE 281 AA; 32137 MW; 8E7E28D16E1105AC CRC64;

Query Match 2.0%; Score 7; DB 16; Length 281;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 104 VFLTWVA 110
Db 186 VFLTWVA 192
|||||

RESULT 39
Q97JM1 PRELIMINARY; PRT; 284 AA.
AC Q97JM1;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SPORULATION PROTEIN IVFB RELATED PROTEIN, PREDICTED
DE METALLOPEPTIDASE.
DE CAC1253.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J.; Berton G.; Onelchenko M.V.; Makarova K.S.; Zeng Q.;
RA Gibson R.; Lee H.M.; Dubois J.; Qiu D.; Hitti J.; Wolf Y.I.;
RA Tatusov R.L.; Sabathe F.; Doucette-Stamm L.; Soucaille P.; Daly M.J.;
RA Bennett G.N.; Koonin E.V.; Smith D.R.;

*Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum."
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL; AE007638; AAK79224.1; -.
DR InterPro; IPR000644; CBS.
DR InterPro; IPR000130; Zn_MTPpeptidse.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 284 AA; 32824 MW; C9CB1436DF7E08FB CRC64;

Query Match 2.0%; Score 7; DB 16; Length 284;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 192 ILFCSEK 198
Db 183 ILFCSEK 189
|||||

RESULT 40
Q9I3I5 PRELIMINARY; PRT; 289 AA.
ID Q9I3I5;
AC Q9I3I5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE CELL DIVISION PROTEIN ZIPA.
GN ZI OR PA1528.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K.; Pham X.-Q.T.; Erwin A.L.; Mizoguchi S.D.; Warrenner P.;
RA Hickey M.J.; Brinkman F.S.L.; Hufnagle W.O.; Kowalik D.J.; Lagrou M.;
RA Garber R.L.; Goltry L.; Tolentino E.; Westbrook-Wadman S.; Yuan Y.;
RA Brody L.L.; Coulter S.N.; Folger K.R.; Kas A.; Larbig K.; Lim R.M.;
RA Smith K.A.; Spencer D.H.; Wong G.K.-S.; Wu Z.; Paulsen I.T.;
RA Reizer J.; Sailer M.H.; Hancock R.E.W.; Lory S.; Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen."
RL Nature 406:959-964(2000).
DR EMBL; AE004581; AAGO4917.1; -.
DR HSSP; P77173; 1F7X.
KW Complete proteome.
SQ SEQUENCE 289 AA; 32236 MW; 720D6FED832B329C CRC64;

Query Match 2.0%; Score 7; DB 16; Length 289;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 LPSVSAR 240
Db 75 LPSVSAR 81
|||||

RESULT 41
Q9IAK3 PRELIMINARY; PRT; 291 AA.
ID Q9IAK3;
AC Q9IAK3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HOMEOBOX TRANSCRIPTION FACTOR SIX3.
GN SIX3.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;

```



```

RN  SEQUENCE FROM N.A.
RX  MEDLINE=20171056; PubMed=10704858;
RA  Zhou X., Hollenmann T., Pieler T., Gruss P.;
RT  "Cloning and expression of xS1x3, the Xenopus homologue of murine
RL  Six3.";
RL  Mech. Dev. 91:327-330(2000).
DR  EMBL: AF183571; AAF63242.1; -.
DR  HSSP: P40424; 1B72.
DR  InterPro: IPR001785; DAHP_synth_1.
DR  InterPro: IPR001356; Homeobox.
DR  Pfam: PF00046; homeobox; 1.
DR  PRINTS: PR00031; HTHREPRESSR.
DR  SMART: SM00389; HOX; 1.
DR  PROSITE: PS50071; HOMEBOX_2; 1.
SQ  SEQUENCE 291 AA; 32780 MW; D7449C85B22A4880 CRC64;

Query Match 2.0%; Score 7; DB 13; Length 291;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 149 YLLENH 155
DB 107 YLLENH 113

RESULT 42
Q908F9 PRELIMINARY; PRT; 300 AA.
ID Q908F9
AC Q908F9
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE M144R.
DE M144R.
GN M144R.
OS Myxoma virus (strain Lausanne).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Leporipoxvirus.
OX NCBI_TaxID=31530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LAUSANNE;
RX MEDLINE=20032073; PubMed=10562494;
RA Cameron C., Hota-Mitchell S., Chen L., Barrett J., Cao J.-X.,
RA Macaulay C., Willier D., Evans D., McFadden G.;
RT "The complete DNA sequence of myxoma virus.";
RT Virology 264:298-318(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=LAUSANNE;
RA Cameron C., Hota-Mitchell S., Chen L., Barrett J., Cao J.-X.,
RA Macaulay C., Willier D., Evans D., McFadden G.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF170726; AAF15032.1; -.
DR HSSP: P10998; 1VVD.
DR InterPro: IPR000436; Sushi_SCR_CCP.
DR Pfam: PF00084; sushi; 3.
DR SMART: SM00032; CCP; 3.
SQ SEQUENCE 300 AA; 34171 MW; F9276A3A390EB099 CRC64;

Query Match 2.0%; Score 7; DB 12; Length 300;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 141 ALVILGT 147
DB 260 ALVILGT 266

RESULT 43
Q9LV43 PRELIMINARY; PRT; 306 AA.
ID Q9LV43

```

```

AC Q9LV43;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE GENOMIC DNA, CHROMOSOME 3, P1 CLONE; MOB24.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=20363099; PubMed=10907853;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by ninety p1,
RT TAC and BAC clones.";
RL DNA Res. 7:217-221(2000).
RL EMBL: AB020746; BAB02012.1; -.
DR InterPro: IPR003439; ABC_transportr.
DR PROSITE: PS00211; ABC_TRANSPORTER; UNKNOWN_1.
SQ SEQUENCE 306 AA; 33465 MW; 76DC343FD2033404 CRC64;

Query Match 2.0%; Score 7; DB 10; Length 306;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 142 LVILGTV 148
DB 119 LVILGTV 125

RESULT 44
Q91ZB7 PRELIMINARY; PRT; 310 AA.
ID Q91ZB7
AC Q91ZB7;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DE G PROTEIN-COUPLED RECEPTOR.
GN MRGE.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RX MEDLINE=21435808; PubMed=11551509;
RA Dong X., Han S.-K., Zylka M.J., Simon M.I., Anderson D.J.;
RT "A Diverse Family of GPCRs Expressed in Specific Subsets of
RT Nociceptive Somatosensory Neurons.";
RL Cell 106:619-632(2001).
DR EMBL: AY042210; AAK91801.1; -.
KW Receptor.
SQ SEQUENCE 310 AA; 34588 MW; 9E44127DCB5E9370 CRC64;

Query Match 2.0%; Score 7; DB 11; Length 310;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 31 LGNGVAL 37
DB 36 LGNGVAL 42

RESULT 45

```

Wed Oct 30 18:40:21 2002

Q9AKA9  
ID Q9AKA9 PRELIMINARY; PRT; 314 AA.  
AC Q9AKA9; 01-JUN-2001 (Tremblrel. 17, Created)  
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE HYPOTHETICAL 34.7 KDA PROTEIN.  
GN RP706.  
OS Rickettsia typhi.  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Rickettsiaceae; Rickettsieae; Rickettsia.  
OX NCBI\_TaxID=785;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=WILMINGTON;  
RX MEDLINE=21219194; PubMed=11319266;  
RA Andersson J.O.; Anderson S.G.E.;  
RT "Pseudogenes, junk DNA and the dynamics of Rickettsia genomes."  
RL Mol. Biol. Evol. 18:829-839(2001).  
DR EMBL; AJ293328; CAC33759.1; -  
DR InterPro; IPR001764; Glyco\_hydro\_3.  
DR Pfam; PF00933; Glyco\_hydro\_3; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 314 AA; 34730 MW; 2E5D188D1AC3E4EC CRC64;  
  
Query Match 2.08; Score 7; DB 2; Length 314;  
Best Local Similarity 100.08; Pred. No. 1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 188 PLGIILF 194  
Db 32 PLGIILF 38  
|||||

Search completed: October 30, 2002, 18:28:31  
Job time : 35 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2002, 18:26:03 ; Search time 13 Seconds  
(without alignments)  
1030.536 Million cell updates/sec

Title: US-09-886-041-2

Perfect score: 346

Sequence: 1 MYNSCCRIEGTISQVMP.....ANSFQSDGQWDPHVEWH 346

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 105224 seqs, 38719550 residues

Word size : 0

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14	4.0	387	1 HM74_HUMAN	P49019 homo sapien
2	8	2.3	340	1 C5AR_GORGO	P79175 gorilla gor
3	8	2.3	340	1 C5AR_MACMU	P79188 macaca mula
4	8	2.3	340	1 C5AR_PANTR	P79240 pan troglod
5	8	2.3	340	1 C5AR_PONPY	P79234 pongo pygma
6	8	2.3	350	1 C5AR_HUMAN	P21730 homo sapien
7	8	2.3	371	1 CML1_MOUSE	P97468 mus musculus
8	8	2.3	371	1 CML1_RAT	O35786 rattus norv
9	8	2.3	373	1 CML1_HUMAN	Q99788 homo sapien
10	7	2.0	131	1 NU3M_CANPA	P48909 candida par
11	7	2.0	141	1 HBAD_CHICK	P02001 gallus gall
12	7	2.0	141	1 HBAD_MELGA	P81024 meleagris g
13	7	2.0	235	1 CLCA_MOUSE	O08585 mus musculus
14	7	2.0	243	1 CLCA_BOVIN	P04973 bos taurus
15	7	2.0	248	1 CLCA_HUMAN	P09496 homo sapien
16	7	2.0	248	1 CLCA_RAT	P08081 rattus norv
17	7	2.0	260	1 IOD2_SHEEP	Q28929 ovis aries
18	7	2.0	264	1 IOD2_RANCA	P49896 rana catesb
19	7	2.0	289	1 ZIPA_PSEAE	Q91315 pseudomonas
20	7	2.0	309	1 YDAA_HAEIN	P44195 haemophilus
21	7	2.0	333	1 CXCI_HUMAN	P46094 homo sapien
22	7	2.0	346	1 FMLR_GORGO	P79176 gorilla gor
23	7	2.0	346	1 FMLR_MACMU	P79189 macaca mula
24	7	2.0	346	1 FMLR_PANTR	P79241 pan troglod
25	7	2.0	346	1 FMLR_PONPY	P79235 pongo pygma
26	7	2.0	350	1 FMLR_HUMAN	P21462 homo sapien
27	7	2.0	355	1 CKR1_MOUSE	P51675 mus musculus
28	7	2.0	360	1 Y572_TREPA	O83582 treponema p
29	7	2.0	363	1 AG22_HUMAN	P50052 homo sapien
30	7	2.0	363	1 AG22_MERUN	Q92026 meriones un
31	7	2.0	363	1 AG22_RAT	P35351 rattus norv
32	7	2.0	364	1 VPAP_HSV7J	P52440 human herpe
33	7	2.0	549	1 SMF2_YEAST	P38778 saccharomyc

RESULT 1

HM74_HUMAN	7	2.0	661	1	OAT2_RAT	O35913 rattus norv
ID	7	2.0	753	1	UL06_HSVEB	P28944 equine herp
AC	7	2.0	1298	1	VGR3_HUMAN	P35916 homo sapien
DT	7	2.0	1363	1	VGR3_MOUSE	P35917 mus musculus
DT	7	2.0	1549	1	TRHY_SHEEP	P22793 ovis aries
DT	6	1.7	50	1	ISPL_GALME	P81905 gallieria me
DT	6	1.7	81	1	NI9M_NEUCR	P42117 neurospora
DT	6	1.7	84	1	LIMU_LIMPO	P02744 limulus pol
DT	6	1.7	91	1	V179_FOWPV	Q9J554 fowlpox vir
DT	6	1.7	100	1	YIUQ_YEAST	P40575 saccharomyc
DT	6	1.7	103	1	TCTA_HUMAN	P57738 homo sapien
DT	6	1.7	103	1	Y273_METJA	Q57721 methanococc
ALIGNMENTS						
HM74_HUMAN STANDARD; PRT; 387 AA.						
AC	P49019;					
DT	01-FEB-1996 (Rel. 33, Created)					
DT	01-FEB-1996 (Rel. 33, Last sequence update)					
DT	16-OCT-2001 (Rel. 40, Last annotation update)					
DE	Probable G protein-coupled receptor HM74.					
GN	HM74.					
OS	Homo sapiens (Human).					
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;					
OC	Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.					
OX	NCBI_TaxID:9606;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	TISSUE=Monocytes;					
RX	MEDLINE=94052629; PubMed=7505609;					
RA	Nomura H., Nielsen B.W., Matsushima K.;					
RT	"Molecular cloning of cDNAs encoding a LD78 receptor and putative					
RT	leukocyte chemotactic peptide receptors.";					
RL	Int. Immunol. 5:1239-1249(1993).					
CC	-I- FUNCTION: ORPHAN RECEPTOR.					
CC	-I- SUBCELLULAR LOCATION: Integral membrane protein.					
CC	-I- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.					
CC	-----					
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration					
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -					
CC	the European Bioinformatics Institute. There are no restrictions on its					
CC	use by non-profit institutions as long as its content is in no way					
CC	modified and this statement is not removed. Usage by and for commercial					
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>					
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).					
CC	-----					
CC	EMBL; D10923; BAA01721.1; -.					
DR	HSSP; P34996; 1DDO.					
DR	GCRD; GCR_0528; -.					
DR	MIM; 606039; -.					
DR	InterPro; IPR000276; GPCR_Rhodpsn.					
DR	Pfam; PF00001; 7tm1; 1.					
DR	PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.					
DR	PROSITE; PS00262; G_PROTEIN_RECEP_F1_2; 1.					
KW	G-protein coupled receptor; Transmembrane.					
FT	DOMAIN 1 28 EXTRACELLULAR (POTENTIAL).					
FT	TRANSMEM 29 50 1 (POTENTIAL).					
FT	DOMAIN 51 63 CYTOPLASMIC (POTENTIAL).					
FT	TRANSMEM 64 85 2 (POTENTIAL).					
FT	DOMAIN 86 102 EXTRACELLULAR (POTENTIAL).					
FT	TRANSMEM 103 123 3 (POTENTIAL).					
FT	DOMAIN 124 142 CYTOPLASMIC (POTENTIAL).					
FT	TRANSMEM 143 163 4 (POTENTIAL).					
FT	DOMAIN 164 194 EXTRACELLULAR (POTENTIAL).					
FT	TRANSMEM 195 209 5 (POTENTIAL).					
FT	DOMAIN 210 236 CYTOPLASMIC (POTENTIAL).					
FT	TRANSMEM 237 256 6 (POTENTIAL).					
FT	DOMAIN 257 273 EXTRACELLULAR (POTENTIAL).					
FT	TRANSMEM 274 298 7 (POTENTIAL).					

FT DOMAIN 299 387 CYTOPLASMIC (POTENTIAL).  
 FT DISULFID 100 177 BY SIMILARITY.  
 SQ SEQUENCE 387 AA; 44481 MW; C24AF562C2343647 CRC64;

Query Match 4.0%; Score 14; DB 1; Length 387;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-06;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 262 ITLSTYNNMMLDP 275  
 Db 278 ITLSTYNNMMLDP 291

## RESULT 2

ID CSAR\_GORGO STANDARD; PRT; 340 AA.  
 AC P79175;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE C5a anaphylatoxin chemotactic receptor (C5a-R) (Fragment).  
 GN C5R1 OR C5AR.  
 OS Gorilla gorilla gorilla (Lowland gorilla).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Gorilla.  
 OX NCBI\_TaxID=9595;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96421539; PubMed=8824156;  
 RA Alvarez V., Coto E., Shen F., Gouzalet-Koces S., Lopez-Larrea C.;  
 RT "Molecular evolution of the N-formyl peptide and C5a receptors in  
 non-human primates."  
 RL Immunogenetics 44:446-452(1996).  
 CC -1- FUNCTION: RECEPTOR FOR THE CHEMOTACTIC AND INFLAMMATORY PEPTIDE  
 ANAPHYLATOXIN C5A. THIS RECEPTOR STIMULATES CHEMOTAXIS, GRANULE  
 ENZYME RELEASE AND SUPEROXIDE ANION PRODUCTION.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- PTM: SULFATION PLAYS A CRITICAL ROLE IN THE ASSOCIATION OF C5AR  
 WITH C5A, BUT NO SIGNIFICANT ROLE IN THE ABILITY OF THE RECEPTOR  
 TO TRANSDUCE A SIGNAL AND MOBILIZE CALCIUM IN RESPONSE TO A SMALL  
 PEPTIDE AGONIST (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X97733; CAA66317.1; .  
 DR GCRDb; GCR\_1102.  
 DR InterPro; IPR000276; GPCR\_Rhodpsn.  
 DR Pfam; PF00001; 7tm\_1; 1.  
 DR PROSITE; PS00237; G\_PROTEIN\_RECF\_F1\_1; 1.  
 DR PROSITE; PS0262; G\_PROTEIN\_RECF\_F1\_2; 1.  
 KW G-protein coupled receptor; Transmembrane; Glycoprotein; Sulfation;  
 Chemotaxis.  
 FT NON\_TER 1 1  
 FT DOMAIN <1 30 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 31 53 1 (POTENTIAL).  
 FT DOMAIN 54 64 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 65 87 2 (POTENTIAL).  
 FT DOMAIN 88 103 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 104 125 3 (POTENTIAL).  
 FT DOMAIN 126 146 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 147 167 4 (POTENTIAL).  
 FT DOMAIN 168 193 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 194 219 5 (POTENTIAL).  
 FT DOMAIN 220 235 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 236 258 6 (POTENTIAL).  
 FT DOMAIN 259 275 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 276 296 7 (POTENTIAL).  
 FT DOMAIN 297 >340 CYTOPLASMIC (POTENTIAL).  
 FT DISULFID 102 181 BY SIMILARITY.  
 FT MOD\_RES 4 4 SULFATION (BY SIMILARITY).  
 FT MOD\_RES 7 7 SULFATION (BY SIMILARITY).  
 FT NON\_TER 340 340  
 SQ SEQUENCE 340 AA; 38235 MW; 214CFCE78A47B304 CRC64;

Query Match 2.3%; Score 8; DB 1; Length 340;  
 Best Local Similarity 100.0%; Pred. No. 2.1;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 56 NLAVADFL 63  
 Db 70 NLAVADFL 77

## RESULT 3

ID CSAR\_MACMU STANDARD; PRT; 340 AA.  
 AC P79188;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE C5a anaphylatoxin chemotactic receptor (C5a-R) (Fragment).  
 GN C5R1 OR C5AR.  
 OS Macaca mulatta (Rhesus macaque).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;  
 OC Cercopitheidae; Macaca.  
 OX NCBI\_TaxID=9544;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96421539; PubMed=8824156;  
 RA Alvarez V., Coto E., Shen F., Gouzalet-Koces S., Lopez-Larrea C.;  
 RT "Molecular evolution of the N-formyl peptide and C5a receptors in  
 non-human primates."  
 RL Immunogenetics 44:446-452(1996).  
 CC -1- FUNCTION: RECEPTOR FOR THE CHEMOTACTIC AND INFLAMMATORY PEPTIDE  
 ANAPHYLATOXIN C5A. THIS RECEPTOR STIMULATES CHEMOTAXIS, GRANULE  
 ENZYME RELEASE AND SUPEROXIDE ANION PRODUCTION.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- PTM: SULFATION PLAYS A CRITICAL ROLE IN THE ASSOCIATION OF C5AR  
 WITH C5A, BUT NO SIGNIFICANT ROLE IN THE ABILITY OF THE RECEPTOR  
 TO TRANSDUCE A SIGNAL AND MOBILIZE CALCIUM IN RESPONSE TO A SMALL  
 PEPTIDE AGONIST (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X97731; CAA66315.1; .  
 DR GCRDb; GCR\_1743.  
 DR InterPro; IPR000276; GPCR\_Rhodpsn.  
 DR Pfam; PF00001; 7tm\_1; 1.  
 DR PROSITE; PS00237; G\_PROTEIN\_RECF\_F1\_1; 1.  
 DR PROSITE; PS0262; G\_PROTEIN\_RECF\_F1\_2; 1.  
 KW G-protein coupled receptor; Transmembrane; Glycoprotein; Sulfation;  
 Chemotaxis.  
 FT NON\_TER 1 1  
 FT DOMAIN <1 30 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 31 53 1 (POTENTIAL).  
 FT DOMAIN 54 64 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 65 87 2 (POTENTIAL).  
 FT DOMAIN 88 103 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 104 125 3 (POTENTIAL).  
 FT DOMAIN 126 146 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 147 167 4 (POTENTIAL).

FT DOMAIN 168 193 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 194 219 5 (POTENTIAL).  
 FT DOMAIN 220 235 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 236 258 6 (POTENTIAL).  
 FT DOMAIN 259 275 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 276 296 7 (POTENTIAL).  
 FT DOMAIN 297 >340 CYTOPLASMIC (POTENTIAL).  
 FT DISULFID 102 181 BY SIMILARITY.  
 FT MOD\_RES 4 4 SULFATION (BY SIMILARITY).  
 FT NON\_TER 7 7 SULFATION (BY SIMILARITY).  
 FT NON\_TER 340 340  
 SQ SEQUENCE 340 AA; 38274 MM; E11F7C73AB97FFBB CRC64;

Query Match 2.3%; Score 8; DB 1; Length 340;  
 Best Local Similarity 100.0%; Pred. No. 2.1;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 NLAVALDFL 63  
 Db 70 NLAVALDFL 77

## RESULT 4

CSAR\_PANTR STANDARD; PRT; 340 AA.  
 AC P79240;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE C5a anaphylatoxin chemotactic receptor (C5a-R) (Fragment).  
 GN CS5R1 OR CS5R.  
 OS Pan troglodytes (Chimpanzee).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pan.  
 OX NCBI\_TaxID=9598;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96421539; PubMed=8824156;  
 RA Alvarez V., Coto E., Sehen F., Gouzalet-Koces S., Lopez-Larrea C.;  
 RT "Molecular evolution of the N-formyl peptide and C5a receptors in  
 non-human primates.";  
 RL Immunogenetics 44:446-452(1996).  
 CC -1- FUNCTION: RECEPTOR FOR THE CHEMOTACTIC AND INFLAMMATORY PEPTIDE  
 ANAPHYLATOXIN C5A. THIS RECEPTOR STIMULATES CHEMOTAXIS, GRANULE  
 ENZYME RELEASE AND SUPEROXIDE ANION PRODUCTION.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- PTM: SULFATION PLAYS A CRITICAL ROLE IN THE ASSOCIATION OF C5AR  
 WITH C5A, BUT NO SIGNIFICANT ROLE IN THE ABILITY OF THE RECEPTOR  
 TO TRANSDUCE A SIGNAL AND MOBILIZE CALCIUM IN RESPONSE TO A SMALL  
 PEPTIDE AGONIST (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X97730; CAA66314.1; -  
 CC GCRDB; GCR\_1551; -  
 CC InterPro; IPR000276; GPCR\_Rhodpsn.  
 CC Pfam; PF00001; 7tm1.1; -  
 CC PROSITE; PS00237; G\_PROTEIN\_RECP\_F1\_1; 1.  
 CC PROSITE; PS0262; G\_PROTEIN\_RECP\_F1\_2; 1.  
 CC G-protein coupled receptor; Transmembrane; Glycoprotein; Sulfation;  
 Chemotaxis.  
 CC NON\_TER 1- 1  
 CC DOMAIN <1 30 EXTRACELLULAR (POTENTIAL).  
 CC TRANSMEM 31 53 1 (POTENTIAL).  
 CC DOMAIN 54 64 CYTOPLASMIC (POTENTIAL).  
 CC TRANSMEM 65 87 2 (POTENTIAL).

FT DOMAIN 88 103 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 104 125 3 (POTENTIAL).  
 FT DOMAIN 126 146 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 147 167 4 (POTENTIAL).  
 FT DOMAIN 168 193 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 194 219 5 (POTENTIAL).  
 FT DOMAIN 220 235 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 236 258 6 (POTENTIAL).  
 FT DOMAIN 259 275 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 276 296 7 (POTENTIAL).  
 FT DOMAIN 297 >340 CYTOPLASMIC (POTENTIAL).  
 FT DISULFID 102 181 BY SIMILARITY.  
 FT MOD\_RES 4 4 SULFATION (BY SIMILARITY).  
 FT NON\_TER 7 7 SULFATION (BY SIMILARITY).  
 FT NON\_TER 340 340  
 SQ SEQUENCE 340 AA; 38209 MM; B3BD4ADBA909513F CRC64;

Query Match 2.3%; Score 8; DB 1; Length 340;  
 Best Local Similarity 100.0%; Pred. No. 2.1;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 NLAVALDFL 63  
 Db 70 NLAVALDFL 77

## RESULT 5

CSAR\_PONPY STANDARD; PRT; 340 AA.  
 AC P79234;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE C5a anaphylatoxin chemotactic receptor (C5a-R) (Fragment).  
 GN CS5R1 OR CS5R.  
 OS Pongo pygmaeus (Orangutan).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pongo.  
 OX NCBI\_TaxID=9600;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96421539; PubMed=8824156;  
 RA Alvarez V., Coto E., Sehen F., Gouzalet-Koces S., Lopez-Larrea C.;  
 RT "Molecular evolution of the N-formyl peptide and C5a receptors in  
 non-human primates.";  
 RL Immunogenetics 44:446-452(1996).  
 CC -1- FUNCTION: RECEPTOR FOR THE CHEMOTACTIC AND INFLAMMATORY PEPTIDE  
 ANAPHYLATOXIN C5A. THIS RECEPTOR STIMULATES CHEMOTAXIS, GRANULE  
 ENZYME RELEASE AND SUPEROXIDE ANION PRODUCTION.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- PTM: SULFATION PLAYS A CRITICAL ROLE IN THE ASSOCIATION OF C5AR  
 WITH C5A, BUT NO SIGNIFICANT ROLE IN THE ABILITY OF THE RECEPTOR  
 TO TRANSDUCE A SIGNAL AND MOBILIZE CALCIUM IN RESPONSE TO A SMALL  
 PEPTIDE AGONIST (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X97732; CAA66316.1; -  
 CC GCRDB; GCR\_1558; -  
 CC InterPro; IPR000276; GPCR\_Rhodpsn.  
 CC Pfam; PF00001; 7tm1.1; -  
 CC PROSITE; PS00237; G\_PROTEIN\_RECP\_F1\_1; 1.  
 CC PROSITE; PS0262; G\_PROTEIN\_RECP\_F1\_2; 1.  
 CC G-protein coupled receptor; Transmembrane; Glycoprotein; Sulfation;  
 Chemotaxis.  
 CC NON\_TER 1 1

FT DOMAIN <1 30 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 31 53 1 (POTENTIAL).  
 FT DOMAIN 54 64 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 65 87 2 (POTENTIAL).  
 FT DOMAIN 88 103 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 104 125 3 (POTENTIAL).  
 FT DOMAIN 126 146 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 147 167 4 (POTENTIAL).  
 FT DOMAIN 168 193 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 194 219 5 (POTENTIAL).  
 FT DOMAIN 220 235 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 236 258 6 (POTENTIAL).  
 FT DOMAIN 259 275 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 276 296 7 (POTENTIAL).  
 FT DOMAIN 297 >340 CYTOPLASMIC (POTENTIAL).  
 FT DISULFID 102 181 BY SIMILARITY.  
 FT MOD\_RES 4 7 SULFATION (BY SIMILARITY).  
 FT NON\_TER 340 340  
 SQ SEQUENCE 340 AA; 38241 MW; 2790316E0C963185 CRC64;

Query Match 2.3%; Score 8; DB 1; Length 340;  
 Best Local Similarity 100.0%; Pred. No. 2.1; Mismatches 0; Indels 0; Gaps 0;  
 Matches 8; Conservative 0;

Qy 56 NLAVADFL 63  
 |||||

Db 70 NLAVADFL 77

## RESULT 6

ID C5AR\_HUMAN STANDARD; PRT; 350 AA.  
 AC P21730;  
 DT 01-MAY-1991 (Rel. 18, Created)  
 DT 01-MAY-1991 (Rel. 18, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE C5a anaphylatoxin chemotactic receptor (C5a-R) (CD88 antigen).  
 GN C5R1 OR C5AR.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91156029; PubMed=1847994;  
 RA Gerard N.P., Gerard C.;  
 RT "The chemotactic receptor for human C5a anaphylatoxin.";  
 RL Nature 349:614-617(1991).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91175748; PubMed=2007135;  
 RA Boulay F., Mery L., Tardif M., Bouchon L., Vignals P.;  
 RT "Expression cloning of a receptor for C5a anaphylatoxin on differentiated HL-60 cells.";  
 RL Biochemistry 30:2993-2999(1991).  
 [3]  
 RP SULFATION.  
 RX MEDLINE=21240629; PubMed=11342590;  
 RA Farzan M., Schnitzler C.E., Vasilieva N., Leung D., Kuhn J.,  
 RA Gerard C., Gerard N.P., Choe H.;  
 RT "Sulfated tyrosines contribute to the formation of the c5a docking site of the human c5a anaphylatoxin receptor.";  
 RL J. Exp. Med. 193:1059-1066(2001).  
 CC -|- FUNCTION: RECEPTOR FOR THE CHEMOTACTIC AND INFLAMMATORY PEPTIDE ANAPHYLATOXIN C5A. THIS RECEPTOR STIMULATES CHEMOTAXIS, GRANULE ENZYME RELEASE AND SUPEROXIDE ANION PRODUCTION.  
 CC -|- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -|- PTM: SULFATION PLAYS A CRITICAL ROLE IN THE ASSOCIATION OF C5AR WITH C5A. BUT NO SIGNIFICANT ROLE IN THE ABILITY OF THE RECEPTOR TO TRANSDUCE A SIGNAL AND MOBILIZE CALCIUM IN RESPONSE TO A SMALL PEPTIDE AGONIST.  
 CC -|- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.

CC -|- DATABASE: NAME=PROW; NOTE=CD guide CD88 entry;  
 CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd88.htm".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; X58674; CAB37830.1; -;  
 DR EMBL; M62505; AA662831.1; -;  
 DR EMBL; X57250; CAA40530.1; -;  
 DR PIR; A37963; A37963;  
 DR PIR; S13646; S13646;  
 DR PIR; S30518; S30518;  
 DR GCRDB; GCR\_0054; -;  
 DR GCRDB; GCR\_0293; -;  
 DR MIN; I13995; -;  
 DR InterPro: IPR000276; GPCR\_Rhodpsn.  
 DR Pfam: PF00001; 7tm\_5; 1;  
 DR PRINTS: PR00237; GPCRHHODPSN.  
 DR PROSITE; PS00237; GLPROTEIN\_RECEP\_FL\_1; 1;  
 DR PROSITE; PS0262; GLPROTEIN\_RECEP\_FL\_2; 1;  
 DR G-protein coupled receptor; Transmembrane; Glycoprotein; Sulfation;  
 KW Chemotaxis.  
 FT DOMAIN 1 37 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 38 60 1 (POTENTIAL).  
 FT DOMAIN 61 71 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 72 94 2 (POTENTIAL).  
 FT DOMAIN 95 110 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 111 132 3 (POTENTIAL).  
 FT DOMAIN 133 153 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 154 174 4 (POTENTIAL).  
 FT DOMAIN 175 200 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 201 226 5 (POTENTIAL).  
 FT DOMAIN 227 242 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 243 265 6 (POTENTIAL).  
 FT DOMAIN 266 282 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 283 303 7 (POTENTIAL).  
 FT DOMAIN 304 350 CYTOPLASMIC (POTENTIAL).  
 FT DISULFID 109 188 BY SIMILARITY.  
 FT MOD\_RES 11 11 SULFATION.  
 FT CARBOHYD 14 14 SULFATION.  
 FT N-LINKED (GLCNAC...) (POTENTIAL).  
 SQ SEQUENCE 350 AA; 39320 MW; DDI0D071D327CDD0 CRC64;  
 Query Match 2.3%; Score 8; DB 1; Length 350;  
 Best Local Similarity 100.0%; Pred. No. 2.2; Mismatches 0; Indels 0; Gaps 0;  
 Matches 8; Conservative 0;  
 Qy 56 NLAVADFL 63  
 |||||  
 Db 77 NLAVADFL 84  
 RESULT 7  
 CMLI\_MOUSE STANDARD; PRT; 371 AA.  
 ID CMLI\_MOUSE  
 AC P97468;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Chemokine receptor like 1 (G-protein coupled receptor DEZ).  
 GN CMLR1 OR DEZ OR GPCR27.  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;

RX MEDLINE-97289630; PubMed-9144535;  
 RA Mether A., Hermey G., Schinke B., Hermans-Borgmeyer I.;  
 RT "A novel G protein-coupled receptor with homology to neuropeptide and  
 RT chemoattractant receptors expressed during bone development.";  
 RL Biochem. Biophys. Res. Commun. 233:336-342(1997).  
 CC -1- FUNCTION: ORPHAN RECEPTOR. COULD BE A CHEMOTACTIC PEPTIDE  
 CC RECEPTOR. MAY HAVE A FUNCTION IN BONE METABOLISM.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING BONE DEVELOPMENT AND IN  
 CC ADULT PARATHYROID GLANDS. EXPRESSED AT DAY E11 IN THE CAUDAL PART  
 CC OF THE TONGUE AND THE UMBILICAL CORD AND THE EXPRESSION IN THE  
 CC TONGUE WAS MAINTAINED THROUGHOUT ADULTHOOD. EXPRESSION INCREASES  
 CC IN BONE AND CARTILAGINOUS FORMING ZONES OF EMBRYO UPTO STAGE E14.5  
 CC AND AT E16.5 EXPRESSION IS SEEN IN THE LUNG.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U79525; AAB53789.1; -;  
 DR GCRDB; GCR\_1301; -;  
 DR MGD; MGI:109603; Cnklr1.  
 DR InterPro; IPR000276; GPCR\_Rhodpsn.  
 DR Pfam; PF00001; 7tm1\_1; 1.  
 DR PRINTS; PR00237; GPCRRHODOPS.  
 DR PROSITE; PS00237; G\_PROTEIN\_RECP\_FL\_1; 1.  
 DR PROSITE; PS0262; G\_PROTEIN\_RECP\_FL\_2; 1.  
 KW G-protein coupled receptor; Transmembrane; Glycoprotein.  
 FT DOMAIN 1 39 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 40 62 1 (POTENTIAL).  
 FT DOMAIN 63 73 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 74 95 2 (POTENTIAL).  
 FT DOMAIN 96 112 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 113 133 3 (POTENTIAL).  
 FT DOMAIN 134 152 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 153 174 4 (POTENTIAL).  
 FT DOMAIN 175 222 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 223 243 5 (POTENTIAL).  
 FT DOMAIN 244 259 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 260 280 6 (POTENTIAL).  
 FT DOMAIN 281 298 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 299 318 7 (POTENTIAL).  
 FT DOMAIN 319 371 CYTOPLASMIC (POTENTIAL).  
 FT CARBOHYD 7 7 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 190 190 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT DISULFID 110 187 POTENTIAL.  
 SQ SEQUENCE 371 AA; 41815 MW; CDBEL19305244C0B3 CRC64;  
  
 Query Match 2.3%; Score 8; DB 1; Length 371;  
 Best Local Similarity 100.0%; Pred. No. 2.3;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 56 NLAVALDFL 63  
 Db 78 NLAVALDFL 85  
 |||||  
  
 RESULT 8  
 CMLL\_RAT STANDARD; PRT; 371 AA.  
 ID CMLL\_RAT  
 AC O35786;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Chemokine receptor-like 1 (G-protein coupled receptor DEZ) (G-protein  
 DE coupled chemoattractant-like receptor).  
 GN CMKLRI OR DEZ.  
 OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ALBINO; TISSUE=Anterior pituitary;  
 RA MEDLINE-98086361; PubMed-9425281;  
 RX Owan C.S.O., Lolait S.J., Santen S., Olde B.;  
 RT "Molecular cloning and tissue distribution of cDNA encoding a novel  
 RT chemoattractant-like receptor.";  
 RL Biochem. Biophys. Res. Commun. 241:390-394(1997).  
 CC -1- FUNCTION: ORPHAN RECEPTOR. COULD BE A CHEMOTACTIC PEPTIDE  
 CC RECEPTOR. MAY HAVE A FUNCTION IN BONE METABOLISM.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- TISSUE SPECIFICITY: HIGH EXPRESSION IN HEART AND LUNG, LOW IN  
 CC SMALL INTESTINES, COLON, KIDNEY, LIVER, UTERUS, AND BRAIN.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AJ002745; CRA05715.1; -;  
 DR GCRDB; GCR\_2583; -;  
 DR InterPro; IPR000276; GPCR\_Rhodpsn.  
 DR Pfam; PF00001; 7tm1\_1; 2.  
 DR PRINTS; PR00237; GPCRRHODOPS.  
 DR PROSITE; PS00237; G\_PROTEIN\_RECP\_FL\_1; 1.  
 DR PROSITE; PS0262; G\_PROTEIN\_RECP\_FL\_2; 1.  
 KW G-protein coupled receptor; Transmembrane; Glycoprotein.  
 FT DOMAIN 1 39 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 40 62 1 (POTENTIAL).  
 FT DOMAIN 63 73 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 74 95 2 (POTENTIAL).  
 FT DOMAIN 96 112 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 113 133 3 (POTENTIAL).  
 FT DOMAIN 134 152 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 153 174 4 (POTENTIAL).  
 FT DOMAIN 175 222 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 223 243 5 (POTENTIAL).  
 FT DOMAIN 244 259 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 260 280 6 (POTENTIAL).  
 FT DOMAIN 281 298 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 299 318 7 (POTENTIAL).  
 FT DOMAIN 319 371 CYTOPLASMIC (POTENTIAL).  
 FT CARBOHYD 7 7 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 190 190 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT DISULFID 110 187 POTENTIAL.  
 SQ SEQUENCE 371 AA; 41722 MW; 3AE308603FEC7801 CRC64;  
  
 Query Match 2.3%; Score 8; DB 1; Length 371;  
 Best Local Similarity 100.0%; Pred. No. 2.3;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 56 NLAVALDFL 63  
 Db 78 NLAVALDFL 85  
 |||||  
  
 RESULT 9  
 CMLL\_HUMAN STANDARD; PRT; 373 AA.  
 ID CMLL\_HUMAN  
 AC Q99788; Q99789; O75748;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Chemokine receptor-like 1 (G-protein coupled receptor DEZ) (G protein-  
 DE coupled receptor ChemR23).  
 GN CMKLRI or DEZ or ChemR23.

OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
RX MEDLINE=97289630; PubMed=9144535;  
RA Methner A., Hermy G., Schinke B., Hermans-Borgmeyer I.;  
RT "A novel G protein-coupled receptor with homology to neuropeptide and  
RT chemoattractant receptors expressed during bone development.";  
RL Biochem. Biophys. Res. Commun. 233:336-342(1997).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM B).  
RX MEDLINE=98264639; PubMed=9603476;  
RA Samson M., Edinger A.L., Stordeur P., Rucker J., Verhasselt V.,  
RA Sharron M., Govaerts C., Mollereau C., Vassart G., Doms R.W.,  
RA Parmentier M.;  
RT "ChemR23, a putative chemoattractant receptor, is expressed in  
RT monocyte-derived dendritic cells and macrophages and is a coreceptor  
RT for HIV and some primary HIV-1 strains.";  
RL Eur. J. Immunol. 28:1689-1700(1998).  
CC -!- FUNCTION: ORPHAN RECEPTOR. COULD BE A CHEMOTACTIC PEPTIDE  
CC RECEPTOR. MAY HAVE A FUNCTION IN BONE METABOLISM. ACTS AS A  
CC CORECEPTOR FOR SEVERAL HIV STRAINS (SIVMAC316, SIVMAC239,  
CC SIVMAC17E-FR AND SIVSM62A), AS WELL AS A PRIMARY HIV-1 STRAIN  
CC (92UG024-2).  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A (SHOWN HERE) AND B; ARE  
CC PRODUCED BY ALTERNATIVE SPLICING.  
CC -!- TISSUE SPECIFICITY: PROMINENTLY EXPRESSED IN DEVELOPING OSSEOUS  
CC AND CARTILAGINOUS TISSUE. ALSO FOUND IN ADULT PARATHYROID GLANDS.  
CC -!- DEVELOPMENTAL STAGE: EXPRESSED DURING BONE DEVELOPMENT.  
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; U79526; AAC51258.1; -;  
DR EMBL; U79527; AAC51259.1; -;  
DR EMBL; Y14838; CAA75112.1; -;  
DR GCRdb; GCR\_1326; -;  
DR GCRdb; GCR\_1327; -;  
DR MIM; 602351; -;  
DR InterPro; IPR000276; GPCR\_Rhodopsn.  
DR PRINTS; PR00001; 7tm\_1; 1.  
DR PROSITE; PR00237; GPCR\_Rhodopsn.  
DR PROSITE; PS00237; G-PROTEIN\_RECP\_FL\_1; 1.  
DR PROSITE; PS0262; G-PROTEIN\_RECP\_FL\_2; 1.  
KW G-protein coupled receptor; Transmembrane; Glycoprotein;  
KW Alternative splicing.  
FT DOMAIN 1 41  
FT DOMAIN 42 64 EXTRACELLULAR (POTENTIAL).  
FT DOMAIN 65 75 1 (POTENTIAL).  
FT DOMAIN 76 97 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 98 114 2 (POTENTIAL).  
FT DOMAIN 115 135 EXTRACELLULAR (POTENTIAL).  
FT DOMAIN 136 154 3 (POTENTIAL).  
FT DOMAIN 155 176 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 177 224 4 (POTENTIAL).  
FT DOMAIN 225 245 EXTRACELLULAR (POTENTIAL).  
FT DOMAIN 246 261 5 (POTENTIAL).  
FT DOMAIN 262 282 6 (POTENTIAL).  
FT DOMAIN 283 300 EXTRACELLULAR (POTENTIAL).  
FT DOMAIN 301 320 7 (POTENTIAL).  
FT DOMAIN 321 373 CYTOPLASMIC (POTENTIAL).  
FT CARBOHYD 9 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 192 192 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT DISULFID 112 189 POTENTIAL.

FT VARSPLIC 1 2 MISSING (IN ISOFORM B).  
FT CONFLICT 248 248 Q -> H (IN REF. 1; AAC51258).  
SQ SEQUENCE 373 AA; 42322 MW; 5244B9738EC3834 CRC64;  
NCBI\_TaxID=9606;  
Query Match 2.3%; Score 8; DB 1; Length 373;  
Best Local Similarity 100.0%; Pred. No. 2.3; Indels 0; Gaps 0;  
Matches 8; Conservative 0; Mismatches 0;  
Qy 56 NLAVADFL 63  
Db 80 NLAVADFL 87  
RESULT 10  
NU3M\_CANPA STANDARD; PRT; 131 AA.  
ID NU3M\_CANPA  
AC P48909;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE NADH-ubiquinone oxidoreductase chain 3 (EC 1.6.5.3).  
GN ND3  
OS Candida parapsilosis (Yeast).  
OG Microchondrion.  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
OX NCBI\_TaxID=5480;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SR2;  
RX MEDLINE=94364940; PubMed=7521869;  
RA Nosek J., Fukuhara H.;  
RT "NADH dehydrogenase subunit genes in the mitochondrial DNA of  
RT yeasts".  
RL J. Bacteriol. 176:5622-5630(1994).  
RN [2]  
RP SEQUENCE OF 46-131 FROM N.A.  
RC STRAIN=SR23;  
RX MEDLINE=95231517; PubMed=7715605;  
RA Nosek J., Dinouel N., Kovac L., Fukuhara H.;  
RT "Linear mitochondrial DNAs from yeasts: telomeres with large tandem  
RT repetitions".  
RL Mol. Gen. Genet. 247:61-72(1995).  
CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.  
CC -!- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 3 FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; X75676; CAA53374.1; -;  
DR EMBL; X76196; CAA53789.1; -;  
DR InterPro; IPR000440; Oxidored\_q4.  
DR Pfam; PF00507; oxidored\_q4; 1.  
KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion.  
SQ SEQUENCE 131 AA; 14682 MW; 5E79B13E66C87572 CRC64;  
Query Match 2.0%; Score 7; DB 1; Length 131;  
Best Local Similarity 100.0%; Pred. No. 10;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 21 LLIAFV 27  
Db 95 LLIAFV 101  
RESULT 11  
HBAD\_CHICK STANDARD; PRT; 141 AA.  
ID HBAD\_CHICK



AC P02001;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Hemoglobin alpha-D chain.  
 GN HBAD.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=WHITE LEGHORN;  
 RX MEDLINE=92020223; PubMed=1656392;  
 RA Lewis W., Lee J.D., Dodgson J.B.;  
 RT "Adult chicken alpha-globin gene expression in transfected QF6 quail  
 RT cells: evidence for a negative regulatory element in the alpha D gene  
 RT region.";  
 RL Nucleic Acids Res. 19:5321-5329(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=83161047; PubMed=6300093;  
 RA Dodgson J.B., Engel J.D.;  
 RT "The nucleotide sequence of the adult chicken alpha-globin genes.";  
 RL J. Biol. Chem. 258:4623-4629(1983).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=82082384; PubMed=6273837;  
 RA Dodgson J.B., McCune K.C., Kusling D.J., Krust A., Engel J.D.;  
 RT "Adult chicken alpha-globin genes alpha A and alpha D: no anemic  
 RT shock alpha-globin exists in domestic chickens.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 78:5998-6002(1981).  
 RN [4]  
 RP PRELIMINARY SEQUENCE.  
 RX MEDLINE=76189993; PubMed=1225908;  
 RA Takei H., Ota Y., Wu K.C., Kiyohara T., Matsuda G.;  
 RT "Amino acid sequence of the alpha chain of chicken AI hemoglobin.";  
 RL J. Biochem. 77:1345-1347(1975).  
 RN [5]  
 RP SEQUENCE OF 1-63 AND 94-122.  
 RC TISSUE=Embryo;  
 RX MEDLINE=82098109; PubMed=7054172;  
 RA Chapman B.S., Hood L.E., Tobin A.J.;  
 RT "Minor early embryonic chick hemoglobin M. Amino acid sequences of  
 RT the epsilon and alpha D chains.";  
 RL J. Biol. Chem. 257:651-658(1982).  
 RN [6]  
 RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).  
 RX MEDLINE=99156923; PubMed=10037733;  
 RA Knapp J.E., Oliveira M.A., Xie Q., Ernst S.R., Riggs A.F.,  
 RA Hackert M.L.;  
 RT "The structural and functional analysis of the hemoglobin D component  
 RT from chicken.";  
 RL J. Biol. Chem. 274:6411-6420(1999).  
 CC -1- FUNCTION: Involved in oxygen transport from the lung to the  
 CC various peripheral tissues.  
 CC -1- SUBUNIT: Heterotetramer of two alpha-D chains and two beta chains.  
 CC The component D forms dimers of tetramers upon deoxygenation.  
 CC -1- TISSUE SPECIFICITY: Red blood cells.  
 CC -1- DEVELOPMENTAL STAGE: In birds, the alpha-D chain occurs in a minor  
 CC hemoglobin component, called hemoglobin d, which is expressed in  
 CC late embryonic and adult life.  
 CC -1- MISCELLANEOUS: REF.5 CHAIN WAS ISOLATED FROM HBM, THE LEAST  
 CC ABUNDANT OF THE FOUR EARLY CHICK HEMOGLOBINS.  
 CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
 CC -----  
 DR EMBL; X59989; CAA42605.1; -  
 DR EMBL; V00411; CAA23702.1; -  
 DR EMBL; M15378; AAA48584.1; ALT\_SEQ.  
 DR EMBL; J00853; AAA48800.1; -  
 DR PIR; A02322; HACHJ.  
 DR PIR; S18672; S18672.  
 DR PDB; 1HBR; 26-MAR-99.  
 DR InterPro; IPR002338; Alpha\_haem.  
 DR InterPro; IPR000971; Globin.  
 DR Pfam; PF00042; globin; 1.  
 DR PRINTS; PR00612; ALPHAHAE.  
 DR PROSITE; PS01033; GLOBIN; 1.  
 KW Heme; Oxygen transport; Transport; Erythrocyte;  
 KW 3D-structure. 58 58 IRON (HEME DISTAL LIGAND).  
 FT METAL 87 87 IRON (HEME PROXIMAL LIGAND).  
 FT CONFLICT 16 16 K -> R (IN REF. 1).  
 FT CONFLICT 107 107 V -> C (IN REF. 3 AND 5).  
 SQ SEQUENCE 141 AA; 15695 MW; 1FE426969B7B5384 CRC64;  
 Query Match 2.0%; Score 7; DB 1; Length 141;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 27 VLGALGN 33  
 DB 62 VLGALGN 68  
 RESULT 12  
 HBAD\_MELGA STANDARD; PRT; 141 AA.  
 AC F81024;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Hemoglobin alpha-D chain.  
 GN HBAD.  
 OS Meleagris gallopavo (Common turkey).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Meleagridae; Meleagris.  
 OX NCBI\_TaxID=9103;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=96066279; PubMed=7576241;  
 RA Eguchi Y., Ikehara T., Kayo S., Eguchi T., Takei H.;  
 RT "Amino acid sequence of alpha- and beta-polypeptide chains of turkey  
 RT (Meleagris gallopavo) hemoglobin.";  
 RL Biol. Chem. Hoppe-Seyler 376:437-440(1995).  
 CC -1- FUNCTION: Involved in oxygen transport from the lung to the  
 CC various peripheral tissues.  
 CC -1- SUBUNIT: Heterotetramer of two alpha-D chains and two beta chains.  
 CC -1- TISSUE SPECIFICITY: Red blood cells.  
 CC -1- DEVELOPMENTAL STAGE: In birds, the alpha-D chain occurs in a minor  
 CC hemoglobin component, called hemoglobin d, which is expressed in  
 CC late embryonic and adult life.  
 CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.  
 CC HSSP; P02001; 1HBR.  
 DR InterPro; IPR002338; Alpha\_haem.  
 DR InterPro; IPR000971; Globin.  
 DR Pfam; PF00042; globin; 1.  
 DR PRINTS; PR00612; ALPHAHAE.  
 DR PROSITE; PS01033; GLOBIN; 1.  
 KW Heme; Oxygen transport; Transport; Erythrocyte.  
 FT METAL 87 87 IRON (HEME DISTAL LIGAND).  
 FT METAL 58 58 IRON (HEME PROXIMAL LIGAND).  
 SQ SEQUENCE 141 AA; 15665 MW; 59942696887954E7 CRC64;  
 Query Match 2.0%; Score 7; DB 1; Length 141;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 27 VLGLGN 33
Db 62 VLGLGN 68

RESULT 13
CLCA_MOUSE
ID CLCA_MOUSE STANDARD; PRT; 235 AA.
AC O08585;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Clathrin light chain A (Lca).
GN CLTA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN SEQUENCE FROM N.A.
RP STRAIN-MDF1.
RA Scott L.M., Mueller L., Collins S.J.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CLATHRIN IS THE MAJOR PROTEIN OF THE POLYHEDRAL COAT OF
CC COATED PITS AND VESICLES.
CC -1- SUBUNIT: CLATHRIN COATS ARE FORMED FROM MOLECULES CONTAINING 3
CC HEAVY CHAINS AND 3 LIGHT CHAINS.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC FACE OF COATED PITS AND
CC VESICLES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U91848; AAB51286.1; -.
DR MGD; MGI:894297; Clta.
DR InterPro; IPR000996; Clathrin_lg_ch.
DR Pfam; PF01086; Clathrin_lg_ch; 1.
DR PROSITE; PS00224; CLATHRIN_LIGHT_CHN_1; FALSE_NEG.
DR PROSITE; PS00581; CLATHRIN_LIGHT_CHN_2; 1.
DR Coated pits; Calcium-binding.
KW DOMAIN 99 161 INVOLVED IN BINDING CLATHRIN HEAVY
FT CHAIN.
SQ SEQUENCE 235 AA; 25557 MW; 886BE97BF27A3E7F CRC64;

Query Match 2.0%; Score 7; DB 1; Length 235;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 30 ALNGVA 36
Db 18 ALNGVA 24

RESULT 14
CLCA_BOVIN
ID CLCA_BOVIN STANDARD; PRT; 243 AA.
AC P04973;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Clathrin light chain A (Lca).
GN CLFA.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;

RN SEQUENCE FROM N.A.
RP MEDLINE=87144634; PubMed=3821891;
RA Jackson A.P., Seow H.-F., Holmes N., Drickamer K., Parham P.;
RT "Clathrin light chains contain brain-specific insertion sequences and
RT a region of homology with intermediate filaments.";
RL Nature 326:154-159(1987).
RN (2)
RP CLATHRIN HEAVY CHAIN-BINDING DOMAIN.
RX MEDLINE=87144642; PubMed=2434865;
RA Brodsky F.M., Galloway C.J., Blank G.S., Jackson A.P.,
RA Seow H.-F., Drickamer K., Parham P.;
RT "Localization of clathrin light-chain sequences mediating heavy-chain
RT binding and coated vesicle diversity.";
RL Nature 326:203-205(1987).
CC -1- FUNCTION: CLATHRIN IS THE MAJOR PROTEIN OF THE POLYHEDRAL COAT OF
CC COATED PITS AND VESICLES.
CC -1- SUBUNIT: CLATHRIN COATS ARE FORMED FROM MOLECULES CONTAINING 3
CC HEAVY CHAINS AND 3 LIGHT CHAINS.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC FACE OF COATED PITS AND
CC VESICLES.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: BRAIN (SHOWN HERE) AND NON-
CC BRAIN; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X04849; CAA28540.1; -.
DR EMBL; X04851; CAA28542.1; -.
DR PIR; A26599; A26599.
DR PIR; B26599; B26599.
DR InterPro; IPR000996; Clathrin_lg_ch.
DR Pfam; PF01086; Clathrin_lg_ch; 1.
DR PROSITE; PS00224; CLATHRIN_LIGHT_CHN_1; 1.
DR PROSITE; PS00581; CLATHRIN_LIGHT_CHN_2; 1.
DR Coated pits; Alternative splicing; Calcium-binding.
KW MOD_RES 1 157 BLOCKED.
FT DOMAIN 95 157 INVOLVED IN BINDING CLATHRIN HEAVY
FT CHAIN.
FT VARSPIC 158 187 MISSING (IN ISOFORM NON-BRAIN).
FT CONFLICT 14 14 P -> H (IN LYMPHOCYTE LCA).
SQ SEQUENCE 243 AA; 26723 MW; B1967A24DC14A06F CRC64;

Query Match 2.0%; Score 7; DB 1; Length 243;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 30 ALNGVA 36
Db 15 ALNGVA 21

RESULT 15
CLCA_HUMAN
ID CLCA_HUMAN STANDARD; PRT; 248 AA.
AC P09496;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Clathrin light chain A (Lca).
GN CLTA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN SEQUENCE FROM N.A.; AND ALTERNATIVE SPLICING.
RP MEDLINE=69034155; PubMed=3267234;

```

RA Jackson A.P., Parham P.;  
 RT "Structure of human clathrin light chains. Conservation of light  
 chain polymorphism in three mammalian species.";  
 RL J. Biol. Chem. 263:16688-16695(1988).  
 [2]  
 RN SEQUENCE FROM N.A. (ISOFORM NON-BRAIN).  
 RP TISSUE=Lung;  
 RC Strausberg R.;  
 RA Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 RL -1- FUNCTION: CLATHRIN IS THE MAJOR PROTEIN OF THE POLYHEDRAL COAT OF  
 CC COATED PITS AND VESICLES.  
 CC -1- SUBUNIT: CLATHRIN COATS ARE FORMED FROM MOLECULES CONTAINING 3  
 CC HEAVY CHAINS AND 3 LIGHT CHAINS.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC FACE OF COATED PITS AND  
 CC VESICLES.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: BRAIN (SHOWN HERE) AND NON-  
 CC BRAIN; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M20471; AAA51817.1; -;  
 DR EMBL; M20472; AAA59505.1; -;  
 DR EMBL; BC009201; AAH09201.1; -;  
 DR PIR; A30752; A30752.  
 DR PIR; A30753; A30753.  
 DR PIR; A31775; A31775.  
 DR MIM; 118960; -;  
 DR InterPro: IPR000996; Clathrin\_lg\_ch.  
 DR Pfam; PF01086; Clathrin\_lg\_ch; 1.  
 DR PROSITE; PS00224; CLATHRIN\_LIGHT\_CHN\_1; 1.  
 DR PROSITE; PS00581; CLATHRIN\_LIGHT\_CHN\_2; 1.  
 CC Coated pits; Alternative splicing; Calcium-binding.  
 FT DOMAIN 100 162 INVOLVED IN BINDING CLATHRIN HEAVY  
 FT CHAIN.  
 FT VARSPLIC 163 192 MISSING (IN ISOFORM NON-BRAIN).  
 SQ SEQUENCE 248 AA; 27076 MW; 8D8A3B49E635D93 CRC64;  
 Query Match 2.0%; Score 7; DB 1; Length 248;  
 Best Local Similarity 100.0%; Pred. No. 18;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 30 ALGNVA 36  
 Db 18 ALGNVA 24  
 RESULT 16  
 CLCA\_RAT  
 ID CLCA\_RAT STANDARD; PRT; 248 AA.  
 AC P08081;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-AUG-1988 (Rel. 08, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Clathrin light chain A (Lca).  
 GN CLTA.  
 OS Rattus norvegicus (Rat).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 CC NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=67178007; PubMed=3563513;  
 RA Kirchhausen T., Scarmato P., Harrison S.C., Monroe J.J., Chow E.P.,  
 RA Mattaliano R.J., Ramchandran K.B., Smart J.E., Ann A.H., Brosius J.;  
 RT "Clathrin light chains LCA and LCB are similar, polymorphic, and  
 RT share repeated heptad motifs.";  
 RL Science 236:320-324(1987).

CC -1- FUNCTION: CLATHRIN IS THE MAJOR PROTEIN OF THE POLYHEDRAL COAT OF  
 CC COATED PITS AND VESICLES.  
 CC -1- SUBUNIT: CLATHRIN COATS ARE FORMED FROM MOLECULES CONTAINING 3  
 CC HEAVY CHAINS AND 3 LIGHT CHAINS.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC FACE OF COATED PITS AND  
 CC VESICLES.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: BRAIN (SHOWN HERE) AND NON-  
 CC BRAIN; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M15882; AAA0868.1; -;  
 DR EMBL; M19261; AAA40870.1; -;  
 DR EMBL; M19260; AAA40869.1; -;  
 DR PIR; A25994; LRR TAL.  
 DR InterPro: IPR000996; Clathrin\_lg\_ch.  
 DR Pfam; PF01086; Clathrin\_lg\_ch; 1.  
 DR PROSITE; PS00224; CLATHRIN\_LIGHT\_CHN\_1; 1.  
 DR PROSITE; PS00581; CLATHRIN\_LIGHT\_CHN\_2; 1.  
 CC Coated pits; Alternative splicing; Calcium-binding.  
 FT DOMAIN 100 162 INVOLVED IN BINDING CLATHRIN HEAVY  
 FT CHAIN.  
 FT VARSPLIC 163 192 MISSING (IN ISOFORM NON-BRAIN).  
 SQ SEQUENCE 248 AA; 26980 MW; C939E85B0FD2E124 CRC64;  
 Query Match 2.0%; Score 7; DB 1; Length 248;  
 Best Local Similarity 100.0%; Pred. No. 18;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 30 ALGNVA 36  
 Db 18 ALGNVA 24  
 RESULT 17  
 AG22\_SHEEP  
 ID AG22\_SHEEP STANDARD; PRT; 260 AA.  
 AC Q28929;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Type-2 angiotensin II receptor (AT2) (Fragment).  
 GN AGTR2.  
 OS Ovis aries (Sheep).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 CC Bovidae; Caprinae; Ovis.  
 CC NCBI\_TaxID=9940;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Kidney;  
 RX MEDLINE=96169644; PubMed=8618791;  
 RA Robillard J.E., Page W.V., Mathews M.S., Schutte B.C., Nuyt A.M.,  
 RA Segar J.L.;  
 RT "Differential gene expression and regulation of renal angiotensin II  
 RT receptor subtypes (AT1 and AT2) during fetal life in sheep.";  
 RL Pediatr. Res. 38:896-904(1995).  
 CC -1- FUNCTION: RECEPTOR FOR ANGIOTENSIN II. MAY HAVE A ROLE IN CELL  
 CC MORPHOGENESIS AND RELATED EVENTS IN GROWTH AND DEVELOPMENT.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

entitles requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

CC EMBL; S81979; AAB36404.1; -.
CC HSSP; P34996; 1DDO.
CC GCRDB; GCR1875; -.
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR Pfam; PF00001; 7tm_1; 1.
DR PROSITE; PS00237; G-PROTEIN_RECF_F1.1; 1.
DR PROSITE; PS0262; G-PROTEIN_RECF_F1.2; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein.
FT NON_TER 1
FT TRANSMEM 1
FT DOMAIN 25 33
FT TRANSMEM 34 55
FT DOMAIN 56 72
FT TRANSMEM 73 93
FT DOMAIN 94 113
FT TRANSMEM 114 132
FT DOMAIN 133 161
FT TRANSMEM 162 187
FT DOMAIN 188 209
FT TRANSMEM 210 231
FT DOMAIN 232 238
FT TRANSMEM 239 >260
FT NON_TER 260 260
SQ SEQUENCE 260 AA; 29549 MW; BB5B48FF029A0B42 CRC64;

Query Match 2.0%; Score 7; DB 1; Length 260;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 FNLAVID 61
DB 37 FNLAVID 43
IIIIIII

RESULT 18
ID IOD2_RANCA STANDARD; PRT; 264 AA.
AC P49896;
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Type II Iodothyronine deiodinase (EC 3.8.1.4) (Type-II 5'deiodinase)
DE (DIOII) (Type 2 DI) (SDII).
GN DIO2 OR TYD12 OR ITD12
OS Rana catesbeiana (Bull frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Ranoides; Ranidae; Rana.
OX NCBI_TaxID=8400;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96070767; PubMed=7592917;
RA Davey J.C., Becker K.B., Schneider M.J., St Germain D.L.,
RA Galton V.A.;
RT "Cloning of a cDNA for the type II Iodothyronine deiodinase.";
RL J. Biol. Chem. 270:26786-26789(1995).
CC -1- FUNCTION: RESPONSIBLE FOR THE DEIODINATION OF T4 (3,5,3',5'-
CC TETRAIODOTHYRONINE) INTO T3 (3,5,3'-TRIODOTHYRONINE) AND OF T3
CC INTO T2 (3,3'-DIODOTHYRONINE). IS RESPONSIBLE FOR THE MAJORITY OF
CC THE INTRACELLULAR T3 IN TISSUES SUCH AS THE PITUITARY, BRAIN AND
CC BROWN FAT BY MEDIATING LOCAL DEIODINATION OF T4, AND IS IMPORTANT
CC IN REGULATING THYROID HORMONE ACTION IN THESE TISSUES. PLAYS A
CC MAJOR ROLE IN DEVELOPMENT.
CC -1- CATALYTIC ACTIVITY: L-thyronine + AH(2) = 3,5,3'-L-triiodo-L-
CC thyronine + iodide + A + H(+).
CC -1- COFACTOR: SELENOCYSTEINE. THE ACTIVE-SITE SELENOCYSTEINE IS
CC ENCODED BY THE OPAL CODON UGA.
CC -1- TISSUE SPECIFICITY: HIGH LEVELS SEEN IN THE METAMORPHOSING TAIL.
CC -1- SIMILARITY: BELONGS TO THE IODOTHYRONINE DEIODINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).


DR EMBL; L42815; AAC42331.2; ALT_SEQ.
DR InterPro; IPR000643; T4_deiodinase.
DR Pfam; PF00837; T4_deiodinase; 1.
DR PROSITE; PS01205; T4_DEIODINASE; 1.
KW Oxidoreductase; Hydrolase; Selenium; Selenocysteine; Transmembrane.
FT TRANSMEM 8 28
FT SE_CYS 124 124
SQ SEQUENCE 264 AA; 29295 MW; DEA617BD35620C88 CRC64;

Query Match 2.0%; Score 7; DB 1; Length 264;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 AVADFL 64
DB 143 AVADFL 149
IIIIIII

RESULT 19
ID ZIPA_PSEAE STANDARD; PRT; 289 AA.
AC Q91315;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Cell division protein zipA homolog.
DE ZIPA OR PA1528.
GN Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 15692 // PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Salier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen."
RL Nature 406:959-964(2000).
CC -1- FUNCTION: Interacts directly with the cell division protein flsZ.
CC Probable receptor for the septal ring structure, may anchor it
CC to the inner-membrane (By similarity).
CC -1- SUBCELLULAR LOCATION: Type IB membrane protein. Inner membrane (By
CC similarity).
CC -1- SIMILARITY: BELONGS TO THE ZIPA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).



DR EMBL; AE004581; AA04917.1; -.
DR HSSP; P77173; 1F7X.
KW Cell division; Separation; Transmembrane; Inner membrane;
FT DOMAIN 1 6
FT TRANSMEM 27
FT DOMAIN 28 289
PERIPLASMIC (POTENTIAL).
CYTOPLASMIC (POTENTIAL).


```

SO SEQUENCE 289 AA; 32236 MW; 720D6FED832B329C CRC64;

Query Match 2.0%; Score 7; DB 1; Length 289;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 234 LPSVSAR 240

Db 75 LPSVSAR 81

RESULT 20

YDAA\_HAEIN STANDARD; PRT; 309 AA.  
AC P44195;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Protein Hil426.  
GN Hil426.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
OX NCBI\_TaxID=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-RD / KW20 / ATCC 51907;  
RX MEDLINE=95350630; PubMed=7542800;  
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
RA Langan H., Takacs B., Evers S., Berndt P., Lahn H.W., Wipf B.,  
RA McKenney K., Sutton G., Fitchugh W., Fields C.A., Gocayne J.D.,  
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudak D.M., Brandon R.C.,  
RA Fine L.D., Fritchman J.L., Fuhrmann J.B., Geoghagen N.S.M.,  
RA Gnehm C.D., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
RA Venter J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus  
influenzae Rd.;"  
RL Science 269:496-512(1995).  
RN [2]  
RP IDENTIFICATION BY MASS SPECTROMETRY.  
RX MEDLINE=20137488; PubMed=10675023;  
RA Langan H., Takacs B., Evers S., Berndt P., Lahn H.W., Wipf B.,  
RA Gray C., Fountoulakis M.;  
RT "Two-dimensional map of the proteome of Haemophilus influenzae.;"  
RL Electrophoresis 21:411-429(2000).  
CC -1- SIMILARITY: BELONGS TO THE UPF0022 (USPA) FAMILY. STRONG, TO  
E.COLI YDAA.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL; U32821; AAC23063.1; ALT\_INIT.

TIGR; Hil426; -.

DR InterPro; IPR000041; Usp.

DR Pfam; PF00582; Usp; 2.

KW Complete proteome.

SO SEQUENCE 309 AA; 35087 MW; C0016A9D91C840A CRC64;

Query Match 2.0%; Score 7; DB 1; Length 309;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 142 LVILGTV 148

Db 268 LVILGTV 274

RESULT 21

CXCL\_HUMAN STANDARD; PRT; 333 AA.  
AC P46094;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Chemokine XC receptor 1 (XC chemokine receptor 1) (Lymphotactin  
receptor) (G protein-coupled receptor GPR5).  
GN CXCR1 OR CXCR1 OR GPR5.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95134353; PubMed=7832990;  
RA Heiber M., Docherty J.M., Shah G., Nguyen T., Cheng R.,  
RA Heng H.H.O., Marchese A., Tsui L.-C., Shi X., George S.R.,  
RA O'Dowd B.F.;  
RT "Isolation of three novel human genes encoding G protein-coupled  
receptors.;"  
RL DNA Cell Biol. 14:25-35(1995).  
RN [2]  
RP CHARACTERIZATION.  
RX MEDLINE=98298178; PubMed=9632725;  
RA Yoshida T., Imai T., Kakizaki M., Nishimura M., Takagi S., Yoshie O.;  
RT "Identification of single C motif-1/lymphotactin receptor XCRL.;"  
RL J. Biol. Chem. 273:16551-16554(1998).  
CC -1- FUNCTION: RECEPTOR FOR CHEMOKINES SCY1 AND SCY2. SUBSEQUENTLY  
CC TRANSDUCES A SIGNAL BY INCREASING THE INTRACELLULAR CALCIUM IONS  
CC LEVEL.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL; L36149; AAA62837.1; -.

DR GCRDB; GCR1842; -.

DR MIM; 600552; -.

DR InterPro; IPR000276; GPCR\_Rhodpsn.

DR Pfam; PF00001; 7tm\_1; 1.

DR PROSITE; PS00237; G\_PROTEIN\_RECEP\_FL1; 1.

DR PROSITE; PS00262; G\_PROTEIN\_RECEP\_FL2; 1.

KW G-protein coupled receptor; Transmembrane.

FT DOMAIN 1 31 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 32 59 1 (POTENTIAL).

FT DOMAIN 60 69 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 70 89 2 (POTENTIAL).

FT DOMAIN 90 103 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 104 125 3 (POTENTIAL).

FT DOMAIN 126 142 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 143 167 4 (POTENTIAL).

FT DOMAIN 168 190 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 191 209 5 (POTENTIAL).

FT DOMAIN 210 225 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 226 250 6 (POTENTIAL).

FT DOMAIN 251 267 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 268 291 7 (POTENTIAL).

FT DOMAIN 292 333 CYTOPLASMIC (POTENTIAL).

FT DISULFID 102 175 BY SIMILARITY.

SO SEQUENCE 333 AA; 36507 MW; C84E7A3EF919CFA5 CRC64;

Query Match 2.0%; Score 7; DB 1; Length 333;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 189 LGIILFC 195      2.0%; Score 7; DB 1; Length 346;
      Best Local Similarity 100.0%; Pred. No. 24;
      Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 198 LGIILFC 204

RESULT 22
FMLR_GORGO
ID FMLR_GORGO STANDARD; PRT; 346 AA.
AC P79176;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE fMet-Leu-Phe receptor (fMLP receptor) (N-formyl peptide receptor)
DE (FPR) (N-formylpeptide chemoattractant receptor) (Fragment).
GN FPR1.
OS Gorilla gorilla gorilla (Lowland gorilla).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Gorilla.
OX NCBI_TaxID=9595;
RN [1]
RP SEQUENCE FROM N.A. PubMed=8824156;
RX MEDLINE=96421539; PubMed=8824156;
RA Alvarez V., Coto E., Sehen F., Gouzalek-Koces S., Lopez-Larrea C.;
RT "Molecular evolution of the N-formyl peptide and c5a receptors in
RT non-human primates.";
RL Immunogenetics 44:446-452(1996).
CC -!- FUNCTION: HIGH AFFINITY RECEPTOR FOR N-FORMYL-METHIONYL PEPTIDES,
CC WHICH ARE POWERFUL NEUTROPHILS CHEMOTACTIC FACTORS. BINDING OF
CC FMLP TO THE RECEPTOR CAUSES ACTIVATION OF NEUTROPHILS. THIS
CC RESPONSE IS MEDIATED VIA A G-PROTEIN THAT ACTIVATES A
CC PHOSPHATIDYLINOSITOL-CALCIUM SECOND MESSENGER SYSTEM.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X97736; CAA66320.1; -
DR GCRDb; GCR_1088; -
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR Pfam; PF00001; 7tm1.1; 1.
DR PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR PROSITE; PS0262; G_PROTEIN_RECEP_F1_2; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein;
KW Chemotaxis.
FT NON_TER 1 1 EXTRACELLULAR (POTENTIAL).
FT DOMAIN <1 24 1 (POTENTIAL).
FT TRANSMEM 25 47 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 48 58 2 (POTENTIAL).
FT TRANSMEM 59 80 3 (POTENTIAL).
FT DOMAIN 81 97 3 (POTENTIAL).
FT TRANSMEM 98 118 4 (POTENTIAL).
FT DOMAIN 119 137 5 (POTENTIAL).
FT TRANSMEM 138 159 6 (POTENTIAL).
FT DOMAIN 160 202 7 (POTENTIAL).
FT TRANSMEM 203 223 7 (POTENTIAL).
FT DOMAIN 224 239 7 (POTENTIAL).
FT TRANSMEM 240 263 7 (POTENTIAL).
FT DOMAIN 264 282 7 (POTENTIAL).
FT TRANSMEM 283 302 7 (POTENTIAL).
FT DOMAIN 303 >346 7 (POTENTIAL).
FT CARBOHYD 1 1 N-LINKED (GLCNAC. .) (POTENTIAL).
FT DISULFID 95 173 N-LINKED (GLCNAC. .) (POTENTIAL).
FT NON_TER 346 346 POTENTIAL.
SQ SEQUENCE 346 AA; 37880 MW; B7FF017586F2D173 CRC64;

QY 56 NLAVADF 62
      2.0%; Score 7; DB 1; Length 346;
      Best Local Similarity 100.0%; Pred. No. 24;
      Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 63 NLAVADF 69

RESULT 23
FMLR_MACMU
ID FMLR_MACMU STANDARD; PRT; 346 AA.
AC P79189;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE fMet-Leu-Phe receptor (fMLP receptor) (N-formyl peptide receptor)
DE (FPR) (N-formylpeptide chemoattractant receptor) (Fragment).
GN FPR1.
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
OX NCBI_TaxID=9544;
RN [1]
RP SEQUENCE FROM N.A. PubMed=8824156;
RX MEDLINE=96421539; PubMed=8824156;
RA Alvarez V., Coto E., Sehen F., Gouzalek-Koces S., Lopez-Larrea C.;
RT "Molecular evolution of the N-formyl peptide and c5a receptors in
RT non-human primates.";
RL Immunogenetics 44:446-452(1996).
CC -!- FUNCTION: HIGH AFFINITY RECEPTOR FOR N-FORMYL-METHIONYL PEPTIDES,
CC WHICH ARE POWERFUL NEUTROPHILS CHEMOTACTIC FACTORS. BINDING OF
CC FMLP TO THE RECEPTOR CAUSES ACTIVATION OF NEUTROPHILS. THIS
CC RESPONSE IS MEDIATED VIA A G-PROTEIN THAT ACTIVATES A
CC PHOSPHATIDYLINOSITOL-CALCIUM SECOND MESSENGER SYSTEM.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X97734; CAA66318.1; -
DR GCRDb; GCR_1741; -
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR Pfam; PF00001; 7tm1.1; 1.
DR PROSITE; PS0262; G_PROTEIN_RECEP_F1_2; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein;
KW Chemotaxis.
FT NON_TER 1 1 EXTRACELLULAR (POTENTIAL).
FT DOMAIN <1 24 1 (POTENTIAL).
FT TRANSMEM 25 47 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 48 58 2 (POTENTIAL).
FT TRANSMEM 59 80 3 (POTENTIAL).
FT DOMAIN 81 97 3 (POTENTIAL).
FT TRANSMEM 98 118 4 (POTENTIAL).
FT DOMAIN 119 137 4 (POTENTIAL).
FT TRANSMEM 138 159 5 (POTENTIAL).
FT DOMAIN 160 202 5 (POTENTIAL).
FT TRANSMEM 203 223 6 (POTENTIAL).
FT DOMAIN 224 239 6 (POTENTIAL).
FT TRANSMEM 240 263 6 (POTENTIAL).
FT DOMAIN 264 282 7 (POTENTIAL).
FT TRANSMEM 283 302 7 (POTENTIAL).
FT DOMAIN 303 >346 7 (POTENTIAL).
FT CARBOHYD 1 1 N-LINKED (GLCNAC. .) (POTENTIAL).
FT DISULFID 95 173 N-LINKED (GLCNAC. .) (POTENTIAL).
FT NON_TER 95 173 POTENTIAL.

```





```

FT DOMAIN 303 >346 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 1 1 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DISULFID 7 7 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 95 173 POTENTIAL.
FT SEQUENCE 346 AA; A012EB86BAA7B315 CRC64;
SQ SEQUENCE 346 AA; A012EB86BAA7B315 CRC64;

Query Match 2.0%; Score 7; DB 1; Length 346;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 56 NLAADF 62
Db 63 NLAADF 69

RESULT 26
FMLR_HUMAN ID FMLR_HUMAN STANDARD; PRT; 350 AA.
AC P21462; Q14939;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE fMet-Leu-Phe receptor (fMLP receptor) (N-formyl peptide receptor)
DE (FPR) (N-formylpeptide chemoattractant receptor).
GN FPR1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxId=9606;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=90267449; PubMed=2161213;
RA Boulay F., Tardif M., Brouchon L., Vignais P.;
RT "Synthesis and use of a novel N-formyl peptide derivative to isolate
RT a human N-formyl peptide receptor cDNA.";
RL Biochem. Biophys. Res. Commun. 168:1103-1109(1990).
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE=91105045; PubMed=2176894;
RA Boulay F., Tardif M., Brouchon L., Vignais P.;
RT "The human N-formylpeptide receptor. Characterization of two cDNA
RT isolates and evidence for a new subfamily of G-protein-coupled
RT receptors.";
RL Biochemistry 29:11123-11133(1990).
RN [3]
RN SEQUENCE FROM N.A.
RX MEDLINE=91286286; PubMed=1712023;
RA Murphy P.M., McDermott D.;
RT "Functional expression of the human formyl peptide receptor in
RT xenopus oocytes requires a complementary human factor.";
RL J. Biol. Chem. 266:12560-12567(1991).
RN [4]
RN SEQUENCE FROM N.A.
RA Perez H.D.;
RL Submitted (MAR-1993) to the EMBL/GenBank/DBJ databases.
RN [5]
RN SEQUENCE FROM N.A.
RX MEDLINE=94040825; PubMed=8224916;
RA Murphy P.M., Tiffany H.L., McDermott D., Ahuja S.K.;
RT "Sequence and organization of the human N-formyl peptide receptor-
RT encoding gene.";
RL Gene 133:285-290(1993).
CC -1- FUNCTION: HIGH AFFINITY RECEPTOR FOR N-FORMYL-METHIONYL PEPTIDES,
CC WHICH ARE POWERFUL NEUTROPHILS CHEMOTACTIC FACTORS. BINDING OF
CC FMLP TO THE RECEPTOR CAUSES ACTIVATION OF NEUTROPHILS. THIS
CC RESPONSE IS MEDIATED VIA A G-PROTEIN THAT ACTIVATES A
CC PHOSPHATIDYLINOSITOL-CALCIUM SECOND MESSENGER SYSTEM.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: NEUTROPHILS
CC -1- POLYMORPHISM: THE VARIANT SEQUENCES ARE THOUGHT TO BE DUE TO
CC ALLELIC VARIATION OF A SINGLE GENE.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.

```

---

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M37128; AAA36362.1; -
CC EMBL; M60626; AAA35846.1; -
CC EMBL; M60627; AAA35847.1; -
CC EMBL; L10820; AAA16863.1; -
CC PIR; A35495; A35495.
CC PIR; A36309; A36309.
CC GCRDB; GCR_0069; -
CC GCRDB; GCR_0070; -
CC GCRDB; GCR_0083; -
CC GCRDB; GCR_0605; -
CC MIM; 136537; -
CC InterPro: IPR000276; GPCR_Rhodpsn.
CC Pfam; PF00001; 7tm1.1; 1.
CC PRINTS; PR00237; GPCRHHODOPSN.
CC PROSITE; PS00237; G_PROTEIN_RECPEP_FL_1; 1.
CC PROSITE; PS0262; G_PROTEIN_RECPEP_FL_2; 1.
CC G-protein coupled receptor; Transmembrane; Glycoprotein;
CC Chemotaxis; Polymorphism.
CC DOMAIN 1 27 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 28 50 1 (POTENTIAL).
CC DOMAIN 51 61 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 62 83 2 (POTENTIAL).
CC DOMAIN 84 100 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 101 121 3 (POTENTIAL).
CC DOMAIN 122 140 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 141 162 4 (POTENTIAL).
CC DOMAIN 163 205 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 206 226 5 (POTENTIAL).
CC DOMAIN 227 242 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 243 266 6 (POTENTIAL).
CC DOMAIN 267 285 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 286 305 7 (POTENTIAL).
CC DOMAIN 306 350 CYTOPLASMIC (POTENTIAL).
CC CARBOHYD 4 4 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 10 10 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC DISULFID 98 176 POTENTIAL.
CC VARIANT 101 101 L->V
CC VARIANT 192 192 /FTid=VAR_003476.
CC VARIANT 346 346 N->K.
CC VARIANT 346 346 /FTid=VAR_003477.
CC CONFLICT 238 238 A->E.
CC CONFLICT 238 238 /FTid=VAR_003478.
CC SEQUENCE 350 AA; 38401 MW; 0663EFF8267E2BD1 CRC64;
R->P (IN REF. 1).

Query Match 2.0%; Score 7; DB 1; Length 350;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 56 NLAADF 62
Db 66 NLAADF 72

RESULT 27
CKRL_MOUSE ID CKRL_MOUSE STANDARD; PRT; 355 AA.
AC P51675;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE C-C chemokine receptor type 1 (C-C CKR-1) (CC-CKR-1) (CCR-1) (CCR1)
DE (Macrophage inflammatory protein-1 alpha receptor) (MIP-1alpha-R)
DE (RANTES-R).

```



```
GN CCR1 OR CMKBR1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV; TISSUE=Peritoneal macrophage;
RX MEDLINE=96072806; PubMed=7594543;
RA Post T.W., Bozic C.R., Rothenberg M.E., Luster A.D., Gerard N.,
RA Gerard C.;
RT "Cloning and differential tissue-specific expression of three mouse
RT beta chemokine receptor-like genes, including the gene for a
RT functional macrophage inflammatory protein-1 alpha receptor.";
RL J. Biol. Chem. 270:17494-17501(1995).
CC -1- FUNCTION: RECEPTOR FOR A C-C TYPE CHEMOKINE. BINDS TO MIP-1-ALPHA,
CC RANTES, AND LESS EFFICIENTLY, TO MIP-1-BETA OR MCP-1 AND
CC SUBSEQUENTLY TRANSDUCES A SIGNAL BY INCREASING THE INTRACELLULAR
CC CALCIUM IONS LEVEL. RESPONSIBLE FOR AFFECTING STEM CELL
CC PROLIFERATION.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: DETECTED IN THE HEART, SPLEEN, LUNG,
CC PERITONEAL EXUDATE CELLS AND LEUKOCYTES.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U93678; AAA86119.1; -.
DR EMBL; U28404; AAA89153.1; -.
DR GCRDB; GCR_1672; -.
DR GCRDB; GCR_1698; -.
DR MGD; MGI:104618; Cmkbr1.
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR Pfam; PF00001; 7tm1_1; 1.
DR PRINTS; PR00237; GPCRHHODOPSN.
DR PROSITE; PS00237; G-PROTEIN_RECP_F1_1; 1.
DR PROSITE; PS00262; G-PROTEIN_RECP_F1_2; 1.
KW G-protein coupled receptor; Transmembrane.
FT DOMAIN 1 34 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 35 60 1 (POTENTIAL).
FT DOMAIN 61 64 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 65 91 2 (POTENTIAL).
FT DOMAIN 92 107 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 108 129 3 (POTENTIAL).
FT DOMAIN 130 146 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 147 171 4 (POTENTIAL).
FT DOMAIN 172 197 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 198 223 5 (POTENTIAL).
FT DOMAIN 224 239 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 240 264 6 (POTENTIAL).
FT DOMAIN 265 281 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 282 305 7 (POTENTIAL).
FT DOMAIN 306 355 CYTOPLASMIC (POTENTIAL).
FT DISULFID 106 183 BY SIMILARITY.
FT CONFLICT 55 55 M -> V (IN REF. 2).
SQ SEQUENCE 355 AA; 40901 MW; FCE9FF7F0E6F38B1 CRC64;

Query Match 2.0%; Score 7; DB 1; Length 355;
Best Local Similarity 100.0%; Pred. No. 24;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 YLFNLAV 59
DB 72 YLFNLAV 78

RESULT 28
Y572_TREPA
ID Y572_TREPA STANDARD; PRT; 360 AA.
AC O83582;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Hypothetical protein TP0572.
GN TP0572.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NICHOLS;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwin M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Artisach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete.";
RL Science 281:375-388(1998).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001232; AAC65551.1; -.
DR TIGR; TP0572; -.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 26 48 POTENTIAL.
FT TRANSMEM 58 80 POTENTIAL.
FT TRANSMEM 89 111 POTENTIAL.
FT TRANSMEM 126 148 POTENTIAL.
FT TRANSMEM 169 191 POTENTIAL.
FT TRANSMEM 195 214 POTENTIAL.
FT TRANSMEM 227 249 POTENTIAL.
SQ SEQUENCE 360 AA; 38633 MW; 798365F156B76BCB CRC64;

Query Match 2.0%; Score 7; DB 1; Length 360;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 VAFVLGA 30
DB 239 VAFVLGA 245

RESULT 29
AG22_HUMAN
ID AG22_HUMAN STANDARD; PRT; 363 AA.
AC P50052; Q13016;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Type-2 angiotensin II receptor (AT2).
GN AGTR2.
```

OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
[1]  
RN SEQUENCE FROM N.A.  
RP TISSUE=Liver;  
RC MEDLINE=95251653; PubMed=7733925;  
RX Martin M.M., Elton T.S.;  
RA "The sequence and genomic organization of the human type 2  
RT angiotensin II receptor.";  
RL Biochem. Biophys. Res. Commun. 209:554-562(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=95309939; PubMed=7790004;  
RA Chassagne C., Beatty B.G., Meloche S.;  
RT "Assignment of the human angiotensin II type 2 receptor gene (AGTR2)  
RL to chromosome Xq22-q23 by fluorescence in situ hybridization.";  
RN Genomics 25:601-603(1995).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Blood;  
RX MEDLINE=95032069; PubMed=7945336;  
RA Kolke G., Horiuchi M., Yamada T., Szpirer C., Jacob H.J., Dzau V.J.;  
RT "Human type 2 angiotensin II receptor gene: cloned, mapped to the X  
RL chromosome, and its mRNA is expressed in the human lung.";  
RN Biochem. Biophys. Res. Commun. 203:1842-1850(1994).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=94242007; PubMed=8185599;  
RA Tsuzuki S., Ichiki T., Nakakubo H., Kitami Y., Guo D.F.,  
RT "Molecular cloning and expression of the gene encoding human  
RL angiotensin II type 2 receptor.";  
RN Biochem. Biophys. Res. Commun. 200:1449-1454(1994).  
RN [5]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lung;  
RX MEDLINE=95091796; PubMed=7999093;  
RA Martin M.M., Su B., Elton T.S.;  
RT "Molecular cloning of the human angiotensin II type 2 receptor cDNA.";  
RL Biochem. Biophys. Res. Commun. 205:645-651(1994).  
RN [6]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=95236034; PubMed=7719706;  
RA Lazard D., Briand-Sutren M.M., Villageois P., Mattei M.-G.,  
RT Strosberg A.D., Nahmias C.;  
RL "Molecular characterization and chromosomal localization of a human  
RN angiotensin II AT2 receptor gene highly expressed in fetal tissues.";  
RN Recept. Channels 2:271-280(1994).  
RN [7]  
RP SEQUENCE OF 1-22 FROM N.A.  
RC TISSUE=Blood;  
RX Katsuya T., Dzau V.J.;  
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.  
RN [8]  
RP SEQUENCE OF 1-16 FROM N.A.  
RC TISSUE=Uterus;  
RX Warnecke C.H., Holzmueister J., Reitz-Zagrosek V., Fleck E.;  
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: RECEPTOR FOR ANGIOTENSIN II. MAY HAVE A ROLE IN CELL  
CC MORPHOGENESIS AND RELATED EVENTS IN GROWTH AND DEVELOPMENT.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -1- TISSUE SPECIFICITY: IN ADULT, HIGHLY EXPRESSED IN MYOMETRIUM WITH  
CC LOWER LEVELS IN ADRENAL GLAND AND FALLOPIAN TUBE. VERY HIGHLY  
CC EXPRESSED IN FETAL KIDNEY AND INTESTINE.  
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed, usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL; U20860; AAA85851.1; -  
DR EMBL; L34579; AAA98990.1; -  
DR EMBL; U10273; AAA61794.1; -  
DR EMBL; U15592; AAA50762.1; -  
DR EMBL; U16957; AAA67753.1; -  
DR EMBL; U27478; AAA84900.1; -  
DR EMBL; X87723; CAA61022.1; -  
DR HSP; P34996; 1DDD.  
DR GCRdb; GCR\_1057; -  
DR GCRdb; GCR\_1245; -  
DR GCRdb; GCR\_1876; -  
DR GCRdb; GCR\_2011; -  
DR GCRdb; GCR\_2027; -  
DR GCRdb; GCR\_2031; -  
DR GCRdb; GCR\_2056; -  
DR MIM; 300034; -  
DR InterPro; IPR000276; GPCR\_Rhodpsn.  
DR Pfam; PF00001; 7tm\_1; 1.  
DR PRINTS; PR00237; GPCRRHODOPS.  
DR PROSITE; PS00237; G-PROTEIN\_RECEP\_FL\_1; 1.  
DR PROSITE; PS00262; G-PROTEIN\_RECEP\_FL\_2; 1.  
DR G-protein coupled receptor; Transmembrane; Glycoprotein; Polymorphism.  
KW DOMAIN 1 45 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 46 71 1 (POTENTIAL).  
FT DOMAIN 72 80 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 81 102 2 (POTENTIAL).  
FT DOMAIN 103 119 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 120 140 3 (POTENTIAL).  
FT DOMAIN 141 160 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 161 179 4 (POTENTIAL).  
FT DOMAIN 180 208 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 209 234 5 (POTENTIAL).  
FT DOMAIN 235 256 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 257 278 6 (POTENTIAL).  
FT DOMAIN 279 285 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 286 313 7 (POTENTIAL).  
FT DOMAIN 314 363 CYTOPLASMIC (POTENTIAL).  
FT CARBOHYD 4 4 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 13 13 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 24 24 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 29 29 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 34 34 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT VARIANT 248 248 R -> K (IN DBSNP:5191).  
FT VARIANT 268 268 C -> W (IN DBSNP:1042860).  
FT CONFLICT 269 269 /FTid=VAR\_011850.  
FT CONFLICT 272 272 W -> C (IN REF. 5).  
FT CONFLICT 272 272 F -> L (IN REF. 4).  
FT CONFLICT 323 323 N -> G (IN REF. 4).  
SQ SEQUENCE 363 AA; 41183 MW; FDD7D4E6F9B43E60 CRC64;  
  
Query Match 2.08; Score 7; DB 1; Length 363;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 55 FNLAVID 61  
Db 84 FNLAVID 90  
|||||  
  
RESULT 30  
AG22 MERUN STANDARD; PRT; 363 AA.  
ID AG22\_MERUN  
AC Q9Z0Z6;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)

```

DE Type-2 angiotensin II receptor (AT2).
GN AGTR2.
OS Meriones unguiculatus (Mongolian jird).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Gerbillinae;
OC Meriones.
OX NCBI_TaxID=10047;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Adrenal gland;
RA Hoe K.-L., Saavedra J.M.;
RT "Cloning and characterization of gerbil AT2.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: RECEPTOR FOR ANGIOTENSIN II. MAY HAVE A ROLE IN CELL
CC MORPHOGENESIS AND RELATED EVENTS IN GROWTH AND DEVELOPMENT (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- PTM: CARBOXYL-TERMINAL SER OR THR RESIDUES MAY BE PHOSPHORYLATED
CC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF080066; ADL9339.1; -
DR InterPro; IPR000276; GPCR_Rhodopsn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PS00237; GPCRHHODOPS.
DR PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR PROSITE; PS0262; G_PROTEIN_RECEP_F1_2; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein;
KW Phosphorylation.
FT DOMAIN 1 45 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 46 71 1 (POTENTIAL).
FT DOMAIN 72 80 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 81 102 2 (POTENTIAL).
FT DOMAIN 103 119 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 120 140 3 (POTENTIAL).
FT DOMAIN 141 160 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 161 179 4 (POTENTIAL).
FT DOMAIN 180 208 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 209 234 5 (POTENTIAL).
FT DOMAIN 235 256 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 257 278 6 (POTENTIAL).
FT DOMAIN 279 285 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 286 313 7 (POTENTIAL).
FT DOMAIN 314 363 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 4 4 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 13 13 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 24 24 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 34 34 N-LINKED (GLCNAC...) (POTENTIAL).
FT MOD_RES 354 354 PHOSPHORYLATION (BY PKC) (POTENTIAL).
SQ SEQUENCE 363 AA; 41435 MW; AOBCA803AD967DBE CRC64;

Query Match 2.0%; Score 7; DB 1; Length 363;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 FNLAVID 61
Dd 84 FNLAVID 90

RESULT 31
AG22_RAT
ID AG22_RAT STANDARD; PRT; 363 AA.
AC P35351;
DT 01-JUN-1994 (Rel. 29, Created)

```

```

DT 01-JUN-1994 (Rel. 29, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE Type-2 angiotensin II receptor (AT2).
GN AGTR2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=94043303; PubMed=8227011;
RA Kambayashi Y., Bardhan S., Takahashi K., Tsuzuki S., Inui H.,
RA Hamakubo T., Inagami T.;
RT "Molecular cloning of a novel angiotensin II receptor isoform
RT involved in phosphotyrosine phosphatase inhibition.";
RL J. Biol. Chem. 268:24543-24546(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Fetal;
RX MEDLINE=94043302; PubMed=8227010;
RA Mukoyama M., Nakajima M., Horiuchi M., Sasamura H., Pratt R.E.,
RA Dzau V.J.;
RT "Expression cloning of type 2 angiotensin II receptor reveals a
RT unique class of seven-transmembrane receptors.";
RL J. Biol. Chem. 268:24539-24542(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=FISCHER; TISSUE=Liver;
RX MEDLINE=96088876; PubMed=7490161;
RA Koike G., Winer E.S., Horiuchi M., Brown D.M., Szpirer C.,
RA Dzau V.J., Jacob H.J.;
RT "Cloning, characterization, and genetic mapping of the rat type 2
RT angiotensin II receptor gene.";
RL Hypertension 26:998-1002(1995).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Liver;
RX MEDLINE=95228708; PubMed=7713098;
RA Inagami T., Iwai N., Sasaki K., Yamano Y., Bardhan S., Chaki S.,
RA Guo D.F., Furuta H., Ohyama K., Kambayashi Y., Takahashi K.,
RA Ichiki T.;
RT "Cloning, expression and regulation of angiotensin II receptors.";
RL Eur. Heart J. 15:104-107(1994).
RN [5]
RP SEQUENCE OF 1-143 FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RX MEDLINE=95322457; PubMed=7599191;
RA Kobayashi S., Ohnishi J., Nibu Y., Nishimatsu S., Umemura S.,
RA Ishii M., Murakami K., Miyazaki H.;
RT "Cloning of the rat angiotensin II type 2 receptor gene and
RT identification of its functional promoter region.";
RL Blochim. Biophys. Acta 1262:155-158(1995).
CC -1- FUNCTION: RECEPTOR FOR ANGIOTENSIN II. MAY HAVE A ROLE IN CELL
CC MORPHOGENESIS AND RELATED EVENTS IN GROWTH AND DEVELOPMENT.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: ABUNDANT EXPRESSION IN FETAL TISSUES, IMMATURE
CC BRAIN, SKIN WOUND AND ATRETIC OVARIAN FOLLICLES.
CC -1- DEVELOPMENTAL STAGE: ABUNDANT IN WHOLE FETUS BUT DECREASES RAPIDLY
CC AFTER BIRTH. IN ADULTS IS HIGHLY EXPRESSED IN THE ADRENAL, PRESENT
CC IN THE BRAIN AND UTERUS BUT UNDETECTABLE IN THE HEART.
CC -1- PTM: CARBOXYL-TERMINAL SER OR THR RESIDUES MAY BE PHOSPHORYLATED.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D16840; BAA04116.1; -
DR EMBL; U01908; AAC52126.1; -
DR EMBL; U22663; AAA86509.1; -

```

```
DR EMBL; D43778; BAA07833.1; -.
DR HSSP; P34996; 1DDD.
DR GCRDb; GCR_0807; -.
DR GCRDb; GCR_0816; -.
DR GCRDb; GCR_1445; -.
DR GCRDb; GCR_1531; -.
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR Pfam; PF00001; 7tm_1; 1.
DR PROSITE; PRO0237; GPCRHHODPSN.
DR PROSITE; PS00237; G_PROTEIN_RECP_FL1; 1.
DR PROSITE; PS0262; G_PROTEIN_RECP_FL2; 1.
KW G-protein coupled receptor; transmembrane; Glycoprotein;
KW Phosphorylation.
FT DOMAIN 1 45 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 46 71 1 (POTENTIAL).
FT DOMAIN 72 80 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 81 102 2 (POTENTIAL).
FT DOMAIN 103 119 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 120 140 3 (POTENTIAL).
FT DOMAIN 141 160 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 161 179 4 (POTENTIAL).
FT DOMAIN 180 208 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 209 234 5 (POTENTIAL).
FT DOMAIN 235 256 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 257 278 6 (POTENTIAL).
FT DOMAIN 279 285 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 286 313 7 (POTENTIAL).
FT DOMAIN 314 363 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 4 4 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 13 13 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 24 24 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 29 29 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 34 34 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT MOD_RES 354 354 PHOSPHORYLATION (BY PKC) (POTENTIAL).
SQ SEQUENCE 363 AA; 41331 MW; 4FE4551A41336650 CRC64;

Query Match 2.0%; Score 7; DB 1; Length 363;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 55 FNLAVID 61
Db 84 FNLAVID 90

RESULT 32
VPAP_HSV7J STANDARD; PRT; 364 AA.
AC P52440;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA polymerase processivity factor (Polymerase accessory protein)
DE (PAP).
GN U27.
OS Human herpesvirus (type 7 / strain JI) (HHV7).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=57278;
RN [1]
RP SEQUENCE FROM N.A.
RA Nicholas J.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ACCESSORY SUBUNIT OF THE DNA POLYMERASE THAT ACTS TO
CC INCREASE THE PROCESSIVITY OF POLYMERIZATION (BY SIMILARITY).
CC -1- SIMILARITY: TO OTHER HERPESVIRUSES POLYMERASE ACCESSORY PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U00062; AAB68900.1; -.
CC PIR; B45154; B45154.
CC PIR; S46731; S46731.
CC SGD; S0001092; SMF2.
CC InterPro; IPR001046; Nrmamp.
CC Pfam; PF01566; Nrmamp; 1.
CC PRINTS; PR00447; NATR5ASCCMP.
CC ProDom; PD001861; Nrmamp; 1.
CC -----

or send an email to license@isb-sib.ch).
CC EMBL; U43400; AAC54689.1; -.
CC DNA-binding; DNA replication.
SQ SEQUENCE 364 AA; 41553 MW; 64520FEDBADA6C37 CRC64;

Query Match 2.0%; Score 7; DB 1; Length 364;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 42 FHMKTWK 48
Db 28 FHMKTWK 34

RESULT 33
SMF2_YEAST STANDARD; PRT; 549 AA.
ID SMF2_YEAST STANDARD; PRT; 549 AA.
AC P38778;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Transporter protein SMF2.
GN SMF2 OR YHR050W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=93077557; PubMed=1447206;
RA West A.H., Clark D.J., Martin J., Neupert W., Hartl F.-U.,
RA Horwich A.L.;
RA "Two related genes encoding extremely hydrophobic proteins suppress a
RA lethal mutation in the yeast mitochondrial processing enhancing
RA protein.";
RL J. Biol. Chem. 267:24625-24633(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=94378003; PubMed=8091229;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA Du Z., Favell A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
RA Latreille P., Louis E.J., Macri C., Macris E., Meneses S., Mouser L.,
RA Nhan M., Rifkin L., Riles L., St Peter H., Trevaskis E., Vaughan K.,
RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
RA Vaudin M.;
RA "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RA VIII.";
RL Science 265:2077-2082(1994).
CC -1- FUNCTION: POSSIBLY INFLUENCE PEP-DEPENDENT PROTEIN IMPORT,
CC POSSIBLY AT THE STEP OF PROTEIN TRANSLOCATION
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
CC (Potential).
CC -1- SIMILARITY: BELONGS TO THE NRMAMP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U00062; AAB68900.1; -.
CC PIR; B45154; B45154.
CC PIR; S46731; S46731.
CC SGD; S0001092; SMF2.
CC InterPro; IPR001046; Nrmamp.
CC Pfam; PF01566; Nrmamp; 1.
CC PRINTS; PR00447; NATR5ASCCMP.
CC ProDom; PD001861; Nrmamp; 1.
```

KW Mitochondrion; Transport; Transmembrane.  
FT TRANSMEM 91 109 POTENTIAL.  
FT TRANSMEM 130 147 POTENTIAL.  
FT TRANSMEM 161 185 POTENTIAL.  
FT TRANSMEM 196 214 POTENTIAL.  
FT TRANSMEM 312 332 POTENTIAL.  
FT TRANSMEM 350 372 POTENTIAL.  
FT TRANSMEM 432 452 POTENTIAL.  
FT TRANSMEM 521 541 POTENTIAL.  
SQ SEQUENCE 549 AA; 59768 MW; 4C200057B82D65E3 CRC64;

Query Match 2.0%; Score 7; DB 1; Length 549;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 133 AGIVCTL 139  
| | | | | | |  
DB 377 AGIVCTL 383

RESULT 34  
OAT2\_RAT STANDARD; PRT; 661 AA.  
AC Q35913; O55224;  
DT 13-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Sodium-independent organic anion transporter 2 (Brain digoxin carrier protein) (Brain-specific organic anion transporter) (OATP-B1).  
DE OATP2.  
GN Rattus norvegicus (Rat).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OC NCBI\_TaxID=10116;  
OX [1]  
RN  
RP SEQUENCE FROM N.A.  
RC STRAIN-SPRAGUE-DAWLEY; TISSUE=Brain;  
RX MEDLINE=97439867; PubMed=9294213;  
RA Noe B., Hagenbuch B., Stieger P.J.;  
RT "Isolation of a multispecific organic anion and cardiac glycoside transporter from rat brain";  
RL Proc. Natl. Acad. Sci. U.S.A. 94:10346-10350(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Retina;  
RX MEDLINE=98380458; PubMed=9712861;  
RA Abe T., Kakyo M., Sakagami H., Tokui T., Nishio T., Tanemoto M., Nomura H., Hebert S.C., Matsuno S., Kondo H., Yawo H.;  
RT "Molecular characterization and tissue distribution of a new organic anion transporter subtype (oatp3) that transports thyroid hormones and taurocholate and comparison with oatp2";  
RL J. Biol. Chem. 273:22395-22401(1998).  
CC -1- FUNCTION: MEDIATES THE NA(+)-INDEPENDENT TRANSPORT OF ORGANIC ANIONS SUCH AS TAUROCHOLATE, CHOLATE, 17BETA-ESTRADIOL-GLUCURONIDE, ESTRONE-3-SULFATE, THE CARDIAC GLYCOSIDES OUABAIN AND DIGOXIN AND THYROID HORMONES. MAY PLAY AN ESPECIALLY IMPORTANT ROLE IN THE BRAIN ACCUMULATION AND TOXICITY OF DIGOXIN AND IN THE HEPATOBIILIARY AND RENAL EXCRETION OF CARDIAC GLYCOSIDES FROM THE BODY.  
CC  
CC  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN BRAIN, LIVER, AND KIDNEY BUT NOT EXPRESSED IN HEART, SPLEEN, LUNG, SKELETAL MUSCLE, AND TESTES.  
CC  
CC  
CC -1- SIMILARITY: BELONGS TO THE SLC21 FAMILY OF TRANSPORTERS.  
CC  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC  
CC

DR EMBL; U88036; AAB80699.1; -  
DR EMBL; U95011; AAC32669.1; -  
DR InterPro: IPR004157; OATP\_C.  
DR InterPro: IPR004156; OATP\_N.  
DR Pfam; PF03137; OATP\_C; 1.  
DR Pfam; PF03132; OATP\_N; 1.  
KW Transmembrane; Transport; Glycoprotein.  
FT TRANSMEM 57 77 POTENTIAL.  
FT TRANSMEM 86 106 POTENTIAL.  
FT TRANSMEM 155 175 POTENTIAL.  
FT TRANSMEM 193 213 POTENTIAL.  
FT TRANSMEM 245 265 POTENTIAL.  
FT TRANSMEM 318 338 POTENTIAL.  
FT TRANSMEM 355 375 POTENTIAL.  
FT TRANSMEM 389 409 POTENTIAL.  
FT TRANSMEM 512 532 POTENTIAL.  
FT TRANSMEM 550 570 POTENTIAL.  
FT TRANSMEM 599 619 POTENTIAL.  
FT CARBOHYD 123 134 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 134 134 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 443 443 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 482 482 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 491 491 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CONFLICT 360 360 Y -> N (IN REF. 2).  
FT CONFLICT 617 617 F -> FF (IN REF. 2).  
SQ SEQUENCE 661 AA; 73251 MW; 692DA69FC85299D2 CRC64;

Query Match 2.0%; Score 7; DB 1; Length 661;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 267 TYMNSML 273  
| | | | | | |  
DB 38 TYMNSML 44

RESULT 35  
UL06\_HSVB STANDARD; PRT; 753 AA.  
AC P28944;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-DEC-1992 (Rel. 24, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE Virion gene 56 protein.  
GN 56.  
OS Equine herpesvirus type 1 (strain Ab4p) (EHV-1).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Alphaherpesvirinae; Varicellovirus.  
OX NCBI\_TaxID=31520;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92295556; PubMed=1318606;  
RA Telford E.A.R., Watson M.S., McBride K., Davison A.J.;  
RT "The DNA sequence of equine herpesvirus-1.";  
RL Virology 189:304-316(1992).  
CC -1- FUNCTION: PRESUMED VIRION PROTEIN; POSSIBLE ROLE IN DNA PACKAGING.  
CC  
CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL6, EHV-1 56, EBV BBF1, HCMV UL104, AND VZV 54.  
CC  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC  
CC EMBL; M86664; AAB02491.1; -  
DR PIR; B36801; WZBEE8.  
DR InterPro: IPR002660; Herpes\_UL6.  
DR Pfam; PF01763; Herpes\_UL6; 1.  
DR ProDom; PD003210; Herpes\_UL6; 1.







"The cDNA-deduced amino acid sequence for trichohyalin, a differentiation marker in the hair follicle, contains a 23 amino acid repeat.";  
RL J. Cell Biol. 110:427-436(1990).  
CC -1- FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY ISODIPEPTIDE BONDS. IT MAY SERVE AS SCAFFOLD PROTEIN, TOGETHER WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL DIFFERENTIATION.  
CC -1- SUBUNIT: HOMODIMER (PROBABLE).  
CC -1- ALTERNATIVE PRODUCTS: AT LEAST TWO FORMS MAY BE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME GENE.  
CC -1- TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING TISSUES SUCH AS THE INNER ROOT SHEATH (IRS) OF HAIR FOLLICLES AND MEDULLA, AND IN THE EPITHELIA OF THE TONGUE, HOOF AND RUMEN.  
CC -1- DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND CALCIUM-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE REPEATS OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STRANDED ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS THE MOST REGULAR AND MAY BIND KIF DIRECTLY BY IONIC INTERACTIONS. DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG DIFFERENT SPECIES.  
CC -1- PTM: KNOWN SUBSTRATE OF TRANSGLUTAMINASE. SOME 200-ARGININES ARE PROBABLY CONVERTED TO CITRULLINES BY PEPTIDYLARGININE DEIMIDASE.  
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE S-100 FAMILY.  
CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; Z18361; CAA79165.1; -.  
DR EMBL; X51695; CAA35992.1; -.  
DR PIR; A34209; A34209.  
DR PIR; S32633; S32633.  
DR PIR; A40691; A40691.  
DR HSP; P02633; 3ICB.  
DR InterPro; IPR002048; EF-hand.  
DR InterPro; IPR001751; S100\_CaBP.  
DR Pfam; PF00036; ehand; 1.  
DR Pfam; PF01023; S\_100; 1.  
DR PROSITE; PS00018; EF\_HAND; 1.  
DR PROSITE; PS00303; S100\_CaBP; FALSE\_NEG.  
DR Repeat; Calcium-binding.  
FT DOMAIN 1 91 S-100 LIKE.  
FT CA\_BIND 22 33 EF-HAND 1 (LOW AFFINITY) (POTENTIAL).  
FT CA\_BIND 62 73 EF-HAND 2 (HIGH AFFINITY) (POTENTIAL).  
FT DOMAIN 413 832 14 X 28 AA APPROXIMATE TANDEM REPEATS.  
FT REPEAT 413 448 1-1.  
FT REPEAT 449 476 1-2.  
FT REPEAT 477 504 1-3.  
FT REPEAT 505 532 1-4.  
FT REPEAT 533 560 1-5.  
FT REPEAT 561 588 1-6.  
FT REPEAT 589 616 1-7.  
FT REPEAT 617 644 1-8.  
FT REPEAT 645 678 1-9.  
FT REPEAT 679 706 1-10.  
FT REPEAT 707 742 1-11.  
FT REPEAT 743 771 1-12.  
FT REPEAT 772 796 1-13.  
FT REPEAT 797 832 1-14.

DOMAIN 1507 938 23 X 23 AA APPROXIMATE TANDEM REPEATS.  
FT REPEAT 938 961 2-1.  
FT REPEAT 962 985 2-2.  
FT REPEAT 986 1021 2-3.  
FT REPEAT 1022 1044 2-4.  
FT REPEAT 1045 1067 2-5.  
FT REPEAT 1068 1090 2-6.  
FT REPEAT 1091 1121 2-7.  
FT REPEAT 1122 1144 2-8.  
FT REPEAT 1145 1167 2-9.  
FT REPEAT 1168 1197 2-10.  
FT REPEAT 1198 1227 2-11.  
FT REPEAT 1228 1250 2-12.  
FT REPEAT 1251 1273 2-13.  
FT REPEAT 1274 1296 2-14.  
FT REPEAT 1297 1319 2-15.  
FT REPEAT 1320 1342 2-16.  
FT REPEAT 1343 1368 2-17.  
FT REPEAT 1369 1391 2-18.  
FT REPEAT 1392 1416 2-19.  
FT REPEAT 1417 1439 2-20.  
FT REPEAT 1440 1461 2-21.  
FT REPEAT 1462 1484 2-22.  
FT REPEAT 1485 1507 2-23.  
FT VARIANT 1145 1197 MISSING (IN SHORT FORM).  
FT VARIANT 1251 1273 MISSING (IN SHORT FORM).  
FT CONFLICT 1399 1599 E -> G (IN REF. 2).  
SQ SEQUENCE 1549 AA; E72FB9FF1326E54E CRC64;  
Query Match 2.0%; Score 7; DB 1; Length 1549;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 203 LRRRQOL 209  
DB 1435 LRRRQOL 1441  
RESULT 39  
ISPL\_GALME  
ID ISPL\_GALME STANDARD; PRT; 50 AA.  
AC P81905;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Inducible serine protease inhibitor 1 (ISPI-1) (Fragment).  
OS Galleria mellonella (Wax moth).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;  
OC Pyraloidea; Pyralidae; Galleriinae; Galleria.  
OX NCBI\_TaxID=71137;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=Hemolymph;  
RX MEDLINE=20193629; Pubmed=10727944;  
RA Froebius A.C., Kanost M.R., Goetz P., Vilcinskas A.;  
RT "Isolation and characterization of novel inducible serine protease inhibitors from larval hemolymph of the greater wax moth Galleria mellonella.";  
RL Eur. J. Biochem. 267:2046-2053(2000).  
CC -1- FUNCTION: INHIBITS TRYPSIN AND THE TOXIN PROTEASE PR2 OF M. ANISOPLIAE. DOES NOT INHIBIT CHYMOTRYPSIN, SUBTILISIN CARLSBERG, PROTEINASE K, PORCINE PANCREATIC ELASTASE AND THE TOXIN PROTEASE PRI OF M. ANISOPLIAE.  
CC -1- DEVELOPMENTAL STAGE: LAST INSTAR LARVAE.  
CC -1- INDUCTION: BY INFECTION.  
KW Serine protease inhibitor.  
FT NON\_TER 50  
SQ SEQUENCE 50 AA; 5368 MW; AD67B6C9D8BC9254 CRC64;  
Query Match 1.7%; Score 6; DB 1; Length 50;  
Best Local Similarity 100.0%; Pred. No. 50;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



```

Qy 128 STRVAA 133
Db 15 STRVAA 20

RESULT 40
N19M_NEUCR
ID N19M_NEUCR STANDARD; PRT; 81 AA.
AC P42117;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE NADH-ubiquinone oxidoreductase 9.5 kDa subunit (EC 1.6.5.3)
DE (EC 1.6.99.3) (Complex I-9.5KD) (Ubiquinone-binding protein).
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariales; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-6.
RC STRAIN=74-OR23-1A;
RX MEDLINE=93075742; PubMed=1445879;
RA Heinrich H., Azevedo J.E., Werner S.;
RT "Characterization of the 9.5-kDa ubiquinone-binding protein of
RT NADH:ubiquinone oxidoreductase (complex I) from Neurospora crassa.";
RL Biochemistry 31:11420-11424(1992).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=93075741; PubMed=1445878;
RA Heinrich H., Werner S.;
RT "Identification of the ubiquinone-binding site of NADH:ubiquinone
RT oxidoreductase (complex I) from Neurospora crassa.";
RL Biochemistry 31:11413-11419(1992).
CC -1- FUNCTION: TRANSFER OF ELECTRONS FROM NADH TO THE RESPIRATORY
CC CHAIN. THE IMMEDIATE ELECTRON ACCEPTOR FOR THE ENZYME IS BELIEVED
CC TO BE UBIQUINONE. THIS SUBUNIT BINDS UBIQUINONE.
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -1- SUBUNIT: COMPLEX I IS COMPOSED OF ABOUT 40 DIFFERENT SUBUNITS.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
CC inner membrane.
CC -1- SIMILARITY: BELONGS TO THE COMPLEX I B9 SUBUNIT FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: S49807; AAB24322.1; -
KW Oxidoreductase; Ubiquinone; NAD; Mitochondrion; Transmembrane.
FT INIT_MET 0 0
FT TRANSMEM 24 42 POTENTIAL.
SQ SEQUENCE 81 AA; 9130 MW; F792B9326AA3CF47 CRC64;

Query Match 1.7%; Score 6; DB 1; Length 81;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 104 VFLTVV 109
Db 37 VFLTVV 42

RESULT 41
LIMU_LIMPO
ID LIMU_LIMPO STANDARD; PRT; 84 AA.
AC P02744;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)

Limulin (Fragment).
OS Limulus polyphemus (Atlantic horseshoe crab).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Merostomata; Xiphosura;
OC Limulidae; Limulus.
OX NCBI_TaxID=6850;
RN [1]
RP SEQUENCE.
RX MEDLINE=78000308; PubMed=409430;
RA Kaplan R., Li S.S.-L., Kehoe J.M.;
RT "Molecular characterization of limulin, a sialic acid binding lectin
RT from the hemolymph of the horseshoe crab, Limulus polyphemus.";
RL Biochemistry 16:4297-4303(1977).
CC -1- FUNCTION: THIS IS A LECTIN THAT BINDS SIALIC ACID.
CC -1- SUBUNIT: HOMOPENTAMER. PENTAXIN (OR PENTAXIN) HAVE A DISCOID
CC ARRANGEMENT OF 5 NONCOVALENTLY BOUND SUBUNITS.
CC -1- PTM: A DISULFIDE BOND LINKS CYS-38 TO A CYS IN THE CARBOXYL HALF
CC OF THE CHAIN OF 163 RESIDUES.
CC -1- PTM: A POSSIBLE CARBOHYDRATE BINDING SITE IS LOCATED BETWEEN
CC POSITIONS 14 AND 20.
CC -1- SIMILARITY: BELONGS TO THE PENTAXIN FAMILY.
DR PIR; A03204; LSHC.
RX HSP; P06205; LIJM.
DR InterPro; IPR001759; Pentaxin.
DR Pfam; PF00334; pentaxin; 1.
DR ProDom; PD002153; Pentaxin; 1.
DR ProSITE; PS00289; PENTAXIN; PARTIAL.
KW Pentaxin; Sialic acid.
FT DOMAIN 1 >84 PENTAXIN.
FT DISULFID 38 ?
FT NON_TER 84 84
SQ SEQUENCE 84 AA; 9529 MW; 00E17AD002B6048B CRC64;

Query Match 1.7%; Score 6; DB 1; Length 84;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 281 SSPSFP 286
Db 16 SSPSFP 21

RESULT 42
V179_FOWPV
ID V179_FOWPV STANDARD; PRT; 91 AA.
AC Q8J554;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Virion envelope protein FPV179.
GN FPV179.
OS Fowlpox virus (FPV).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Avipoxvirus.
OX NCBI_TaxID=10261;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20193820; PubMed=10729156;
RA Afonso C.L., Tulman E.R., Lu Z., Zsak L., Kutish G.F., Rock D.L.;
RT "The genome of fowlpox virus.";
RL J. Virol. 74:3815-3831(2000).
CC -1- SIMILARITY: BELONGS TO THE POXVIRUSES A14 FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF198100; AAF44523.1; -
KW Envelope protein.
SQ SEQUENCE 91 AA; 9950 MW; 64FE7AEECFD3849D CRC64;

```

```
Query Match 1.7%; Score 6; DB 1; Length 91;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 189 LGIILF 194
DB 60 LGIILF 65

RESULT 43
YIUO_YEAST STANDARD; PRT; 100 AA.
AC P40575;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE Hypothetical 11.7 kDa protein in MAL5-MRS1 intergenic region.
GN YIR020C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Barrall B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z38061; CAA86177.1; -
DR PIR; S48479; S48479.
DR SGD; S0001459; YIR020C.
KW Hypothetical protein.
SQ SEQUENCE 100 AA; 11672 MW; EA0A3C0AF0124218 CRC64;

Query Match 1.7%; Score 6; DB 1; Length 100;
Best Local Similarity 100.0%; Pred. No. 90;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 104 VPLTVV 109
DB 35 VPLTVV 40

RESULT 44
TCCTA_HUMAN STANDARD; PRT; 103 AA.
AC P57738;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE T-cell leukemia translocation-associated gene protein.
GN TCCTA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=95246031; PubMed=7728759;
RA Aplan P.D., Johnson B.E., Russell E., Chervinsky D.S., Kirsch I.R.;
RA "Cloning and characterization of TCCTA, a gene located at the site of a
RA t(1;3) translocation.";
RA Cancer Res. 55:1917-1921(1995).
CC -!- TISSUE SPECIFICITY: UBIQUITOUS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L41143; AAA97514.1; -
DR MIM; 600690; -
SQ SEQUENCE 103 AA; 11341 MW; 62FA7ECC2241D85E CRC64;

Query Match 1.7%; Score 6; DB 1; Length 103;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 VLGAIG 32
DB 17 VLGAIG 22

RESULT 45
Y273_METJA STANDARD; PRT; 103 AA.
ID Y273_METJA
AC Q57721;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ0273.
GN Methanococcus jannaschii.
OS Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kellavag A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
RA jannaschii.";
RA Science 273:1058-1073(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U67482; AAB98261.1; -
DR TIGR; MJ0273; -
SQ SEQUENCE 103 AA; 11372 MW; BDE81755780AF7EB CRC64;

Query Match 1.7%; Score 6; DB 1; Length 103;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Oy 61 DFLMI 66  
| | | | |  
Db 46 DFLMI 51

Search completed: October 30, 2002, 18:26:29  
Job time : 17 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2002, 18:26:03 ; Search time 18 Seconds

(without alignments)  
1847.051 Million cell updates/sec

Title: US-09-886-041-2

Perfect score: 346

Sequence: 1 MYNGSCCRIEGDTISQVMP.....ANSFQSQDQWDPHVEWH 346

Scoring table:

OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 283138 seqs, 96089334 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

PIR\_71:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14	4.0	387	2 I69202	G protein-coupled
2	9	2.6	391	2 T32517	hypothetical prote
3	8	2.3	200	2 AC0343	probable CDP-alcoh
4	8	2.3	350	1 A37963	complement C5a ana
5	8	2.3	371	2 JC5796	probable chemoattr
6	8	2.3	371	2 JC5498	G protein-coupled
7	8	2.3	438	2 AH0031	proton glutamate s
8	8	2.3	572	2 T34345	hypothetical prote
9	7	2.0	83	2 T23070	hypothetical prote
10	7	2.0	86	2 S54752	NADH dehydrogenase
11	7	2.0	120	2 F83244	hypothetical prote
12	7	2.0	141	1 HACH1	hemoglobin alpha-D
13	7	2.0	141	2 S56103	hemoglobin alpha-D
14	7	2.0	172	2 G64555	hypothetical prote
15	7	2.0	172	2 A71953	hypothetical prote
16	7	2.0	206	2 F75353	hypothetical prote
17	7	2.0	237	2 AG3474	carbamoyl-phosphat
18	7	2.0	243	2 A26599	clathrin light cha
19	7	2.0	248	2 A31775	clathrin light cha
20	7	2.0	258	2 AH0360	probable dimethyl
21	7	2.0	264	2 T10530	thyroxine delodina
22	7	2.0	272	2 T33515	hypothetical prote
23	7	2.0	281	2 C82547	hypothetical prote
24	7	2.0	284	2 E97054	sporulation protei
25	7	2.0	286	1 LRRTA1	clathrin light cha
26	7	2.0	289	2 E83454	cell division prot
27	7	2.0	314	2 E71630	hypothetical prote
28	7	2.0	326	2 AB3607	3-oxoacyl-(acyl-ca
29	7	2.0	329	2 AI2023	protochlorophyllid

30	7	2.0	333	2 I65989	G protein-coupled
31	7	2.0	340	2 G64029	hypothetical prote
32	7	2.0	350	2 A42009	N-formyl peptide r
33	7	2.0	355	2 I49339	macrophage inflam
34	7	2.0	360	2 B71309	hypothetical prote
35	7	2.0	363	2 A49092	angiotensin II rec
36	7	2.0	363	2 JC2543	angiotensin II rec
37	7	2.0	364	2 T41929	polymerase process
38	7	2.0	395	2 B83313	probable type II s
39	7	2.0	400	2 T49810	hypothetical prote
40	7	2.0	414	2 A70214	hypothetical prote
41	7	2.0	422	2 AG1804	sugar transport pr
42	7	2.0	422	2 AI1430	sugar transport pr
43	7	2.0	448	2 T27955	hypothetical prote
44	7	2.0	452	2 AH0708	PTS system, cellob
45	7	2.0	473	2 T24686	hypothetical prote

ALIGNMENTS

RESULT 1

I69202

G protein-coupled receptor HM74 - human

C:Species: Homo sapiens (man)

C>Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 19-May-2000

C:Accession: I69202

R:Nomura, H.; Nielsen, B.W.; Matsushima, K.

Int. Immunol. 5, 1239-1249, 1993

A:Title: Molecular cloning of cDNAs encoding a LD78 receptor and putative leukocyte c

A:Reference number: I54751; MUID:94092629

A:Accession: I69202

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A:Residues: 1-387 <RES>

A:Cross-references: GB:D10923; NID:G219866; PIDN:BAA01721.1; PID:G219867

C:Genetics:

A:Gene: HM74

C:Superfamily: G protein-coupled receptor 4

Query Match 4.0%; Score 14; DB 2; Length 387;

Best Local Similarity 100.0%; Pred. No. 3.7e-06;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 262 ITLSFTYMSMLDP 275

Db 278 ITLSFTYMSMLDP 291

RESULT 2

T32517

hypothetical protein C49A9.7 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 20-Apr-2000

C:Accession: T32517

R:Fulton, B.; Wohldmann, P.

submitted to the EMBL Data Library, December 1997

A:Description: The sequence of C. elegans cosmid C49A9.

A:Reference number: Z21184

A:Status: preliminary;

A:Accession: T32517

A:Molecule type: DNA

A:Residues: 1-391 <FUL>

A:Cross-references: EMBL:AF036693; PIDN:AAB88331.1; GSPDB:GN00022; CESP:C49A9.7

A:Experimental source: strain Bristol N2; clone C49A9

C:Genetics:

A:Gene: CESP:C49A9.7

A:Map position: 4

A:Introns: 21/3; 84/2; 128/2; 159/2; 199/3; 233/2; 263/3; 347/2

C:Superfamily: neurokinin 1 receptor

Query Match

Best Local Similarity 2.6%; Score 9; DB 2; Length 391;

Mismatches 0; Indels 0; Gaps 0;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 53 YLFNLAVAD 61  
| | | | | | | | | |  
Db 99 YLFNLAVAD 107

RESULT 3  
AC0343  
C:Species: Yersinia pestis  
C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 02-Nov-2001  
C:Accession: AC0343  
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; Hill, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall, Nature 413, 523-527, 2001  
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A:Reference number: AB0001; MUID:21470413; PMID:11586360  
A:Accession: AC0343  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-200 <KUR>  
A:Cross-references: GB:AL590842; PIDN:CAC93050.1; PID:gl5980788; GSPDB:GN00175  
C:Genetics:  
A:Gene: YPO2817

Query Match 2.3%; Score 8; DB 2; Length 200;  
Best Local Similarity 100.0%; Pred. No. 3.5;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 25 AFVLGALG 32  
| | | | | | | | | |  
Db 152 AFVLGALG 159

RESULT 4  
A37963  
C:Species: Homo sapiens (man)  
C:Date: 22-Jan-1993 #sequence\_revision 14-Feb-1997 #text\_change 21-Jul-2000  
C:Accession: A37963; S13646; I52417; S30518  
R:Boulay, F.; Mery, L.; Tardif, M.; Brouchon, L.; Vignais, P. Biochemistry 30, 2993-2999, 1991  
A:Title: Expression cloning of a receptor for C5a anaphylatoxin on differentiated HL-60  
A:Reference number: A37963; MUID:91175748  
A:Accession: A37963  
A:Molecule type: mRNA  
A:Residues: 1-350 <BOU>  
A:Cross-references: GB:J05327; NID:gl79699; PIDN:AAA62831.1; PID:gl79700  
R:Gerard, N.P.; Gerard, C. Nature 349, 614-617, 1991  
A:Title: The chemotactic receptor for human C5a anaphylatoxin.  
A:Reference number: S13646; MUID:91156029  
A:Accession: S13646  
A:Molecule type: mRNA  
A:Residues: 1-350 <GER>  
A:Cross-references: EMBL:X59674; NID:g29568; PIDN:CAB37830.1; PID:g4467832  
R:Gerard, N.P.; Bao, L.; Xiao-Ping, H.; Eddy, R.L. Biochemistry 32, 1243-1250, 1993  
A:Title: Human chemotaxis receptor genes cluster at 19q13.3-13.4. Characterization of the  
A:Reference number: I52417; MUID:93192225  
A:Accession: I52417  
A:Status: translated from GB/ENBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-3 <RES>  
A:Cross-references: GB:S56556; GB:S56557; NID:g298577; NID:g298578  
C:Genetics:  
A:Gene: GDB:C5R1; C5A; C5AR  
A:Cross-references: GDB:I28856; OMIM:113995  
A:Map position: 19q13.3-19q13.4  
A:Introns: 1/3  
A:Note: the list of introns may be incomplete

C:Function:  
A:Description: mediates the inflammatory and chemotactic responses of polymorphonucle  
n  
C:Superfamily: vertebrate rhodopsin  
C:Keywords: chemotaxis; G protein-coupled receptor; glycoprotein; inflammation; polym  
F:1-37/Domain: extracellular #status predicted <EX1>  
F:38-61/Domain: transmembrane #status predicted <TM1>  
F:62-71/Domain: intracellular #status predicted <IN1>  
F:72-94/Domain: transmembrane #status predicted <TM2>  
F:95-110/Domain: extracellular #status predicted <EX2>  
F:111-132/Domain: transmembrane #status predicted <TM3>  
F:133-149/Domain: intracellular #status predicted <IN2>  
F:150-174/Domain: transmembrane #status predicted <TM4>  
F:175-206/Domain: extracellular #status predicted <EX3>  
F:207-227/Domain: transmembrane #status predicted <TM5>  
F:228-242/Domain: intracellular #status predicted <IN3>  
F:243-264/Domain: transmembrane #status predicted <TM6>  
F:265-283/Domain: extracellular #status predicted <EX4>  
F:284-307/Domain: transmembrane #status predicted <TM7>  
F:308-350/Domain: intracellular #status predicted <IN4>  
F:5/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 2.3%; Score 8; DB 1; Length 350;  
Best Local Similarity 100.0%; Pred. No. 5.7;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 56 NLAVADFL 63  
| | | | | | | | | |  
Db 77 NLAVADFL 84

RESULT 5  
JC5796  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 24-Jan-1998 #sequence\_revision 13-Mar-1998 #text\_change 20-Jun-2000  
C:Accession: JC5796  
R:Owman, C.; Lolait, S.J.; Santen, S.; Olde, B. Biochem. Biophys. Res. Commun. 241, 390-394, 1997  
A:Title: Molecular cloning and tissue distribution of a novel chemoattr  
A:Reference number: JC5796; MUID:98086361  
A:Accession: JC5796  
A:Molecule type: mRNA  
A:Residues: 1-371 <OMW>  
A:Cross-references: DBJ:AJ002745; NID:g2624397; PIDN:CAA05715.1; PID:g2624398  
C:Experimental source: liver  
C:Comment: This protein regulates the trafficking of immune cells during a microbial  
C:Superfamily: vertebrate rhodopsin  
C:Keywords: glycoprotein; receptor; transmembrane protein  
F:39-62/Domain: transmembrane #status predicted <TM1>  
F:74-94/Domain: transmembrane #status predicted <TM2>  
F:112-135/Domain: transmembrane #status predicted <TM3>  
F:156-177/Domain: transmembrane #status predicted <TM4>  
F:206-237/Domain: transmembrane #status predicted <TM5>  
F:259-280/Domain: transmembrane #status predicted <TM6>  
F:302-317/Domain: transmembrane #status predicted <TM7>  
F:7.190/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 2.3%; Score 8; DB 2; Length 371;  
Best Local Similarity 100.0%; Pred. No. 6;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 56 NLAVADFL 63  
| | | | | | | | | |  
Db 78 NLAVADFL 85

RESULT 6  
JC5498  
C:Species: Mus musculus (house mouse)  
C:Date: 07-Jul-1997 #sequence\_revision 29-Aug-1997 #text\_change 19-May-2000  
C:Accession: JC5498

R;Methner, A.; Hermey, G.; Schinke, B.; Hermans-Borgmeyer, I.  
Biochem. Biophys. Res. Commun. 233, 336-342, 1997  
A;Title: A novel G protein-coupled receptor with homology to neuropeptide and chemoattractant  
A;Reference number: JC5498; MUID:97289630  
A;Contents: Brain  
A;Accession: JC5498  
A;Molecule type: mRNA  
A;Residues: 1-371 <MET>  
A;Cross-references: GB:U79525; NID:gl732346; PIDN:AAB53789.1; PID:gl732347  
C;Comment: This protein is involved in the bone metabolism.  
C;Superfamily: vertebrate rhodopsin  
F:110-187/Disulfide bonds: #status predicted

Query Match 2.3%; Score 8; DB 2; Length 371;  
Best Local Similarity 100.0%; Pred. No. 6;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 NLAVADFL 63  
Db 78 NLAVADFL 85  
|||||

RESULT 7  
AH0031  
proton glutamate symport protein [imported] - Yersinia pestis (strain CO92)  
C;Species: Yersinia pestis  
C;Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Nov-2001  
C;Accession: AH0031  
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; Ill, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrrell, Nature 413, 523-527, 2001  
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A;Reference number: AB0001; MUID:21470413; PMID:11586360  
A;Accession: AH0031  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-438 <RUR>  
A;Cross-references: GB:AL590842; PIDN:CAC89115.1; PID:gl5978353; GSPDB:GN00175  
C;Genetics: gltP  
A;Gene: gltP  
C;Superfamily: C4-dicarboxylate carrier protein

Query Match 2.3%; Score 8; DB 2; Length 438;  
Best Local Similarity 100.0%; Pred. No. 6.9;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 141 ALVILGTV 148  
Db 245 ALVILGTV 252  
|||||

RESULT 8  
T34345  
hypothetical protein T12A2.15 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C;Accession: T34345  
R;Latreille, P.  
submitted to the EMBL Data Library, March 1996  
A;Description: The sequence of C. elegans cosmid T12A2.  
A;Reference number: Z21511  
A;Accession: T34345  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-572 <LAT>  
A;Cross-references: EMBL:U13019; PIDN:AAC24452.1; GSPDB:GN00021; CESP:T12A2.15  
A;Experimental source: strain Bristol N2; clone T12A2  
C;Genetics:  
A;Gene: CESP:T12A2.15  
A;Map position: 3  
A;Introns: 53/1; 82/2; 147/3; 215/3; 274/3; 304/3; 405/3; 465/3; 518/3

Query Match 2.3%; Score 8; DB 2; Length 572;  
Best Local Similarity 100.0%; Pred. No. 8.7;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 GDIPCRVG 91  
Db 165 GDIPCRVG 172  
|||||

RESULT 9  
T23070  
hypothetical protein H12D21.3 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
C;Accession: T23070; T27558  
R;McMurray, A.  
submitted to the EMBL Data Library, March 1997  
A;Reference number: Z19671  
A;Accession: T23070  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-83 <WIL>  
A;Cross-references: EMBL:292849; PIDN:CAB07424.1; GSPDB:GN00023; CESP:H12D21.3  
A;Experimental source: clone H12D21  
R;Ainscough, R.  
submitted to the EMBL Data Library, August 1996  
A;Reference number: Z20387  
A;Accession: T27558  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-83 <WI2>  
A;Cross-references: EMBL:278067; PIDN:CAB01527.1; GSPDB:GN00023; CESP:ZC412.8  
A;Experimental source: clone ZC412  
C;Genetics:  
A;Gene: CESP:H12D21.3; CESP:ZC412.8  
A;Map position: 5

Query Match 2.0%; Score 7; DB 2; Length 83;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 LLIVAFV 27  
Db 10 LLIVAFV 16  
|||||

RESULT 10  
S54752  
NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 3 - yeast (Candida parapsilosis) m  
C;Species: mitochondrion Candida parapsilosis  
C;Date: 27-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 07-Dec-1999  
C;Accession: S54752; S44276  
R;Nosek, J.; Dinouel, N.; Kovac, L.; Fukuhara, H.  
Mol. Gen. Genet. 247, 61-72, 1995  
A;Title: Linear mitochondrial DNAs from yeasts: telomeres with large tandem repetitive  
A;Reference number: S54752; MUID:95231517  
A;Accession: S54752  
A;Status: translation not shown  
A;Molecule type: DNA  
A;Residues: 1-86 <NOS>  
A;Cross-references: EMBL:X76196; NID:9479138; PIDN:CAA53789.1; PID:9479139  
C;Genetics:  
A;Gene: ND3  
A;Genome: mitochondrion  
A;Genetic code: SGC3  
C;Superfamily: NADH dehydrogenase (ubiquinone) chain 3  
C;Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation

Query Match 2.0%; Score 7; DB 2; Length 86;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 LLIVAFV 27

Db 50 LLIVAFV 56  
|||||||

## RESULT 11

F83244 hypothetical protein PA3203 [Imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: F83244  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Boman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen  
A:Reference number: A82950; MUID:20437337  
A:Accession: F83244  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-120 <STO>  
A:Cross-references: GB:AE004744; GB:AE004091; NID:9949317; PIDN:AAG06591.1; GSPDB:GN001  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA3203

Query Match 2.0%; Score 7; DB 2; Length 120;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 25 AFVLGAL 31  
|||||||

Db 101 AFVLGAL 107

## RESULT 12

HACH1 hemoglobin alpha-D chain - chicken  
C:Species: Gallus gallus (chicken)  
C:Date: 24-Apr-1984 #sequence\_revision 20-Sep-1984 #text\_change 03-Mar-2000  
C:Accession: B92421; B93892; A91937; B92377; S18672; A02322  
R:Dodgson, J.B.; Engel, J.D.  
J. Biol. Chem. 258, 4623-4629, 1983  
A:Title: The nucleotide sequence of the adult chicken alpha-globin genes.  
A:Reference number: A92421; MUID:83161047  
A:Accession: B92421  
A:Molecule type: DNA  
A:Residues: 1-141 <DOD>  
A:Cross-references: GB:J00853; NID:g211866; PIDN:AAA4800.1; PID:g211869  
A:Note: The gene from which this sequence was translated is unusual in that the second intron is located within the coding sequence.  
R:Dodgson, J.B.; McCune, K.C.; Rusling, D.J.; Krust, A.; Engel, J.D.  
Proc. Natl. Acad. Sci. U.S.A. 78, 5998-6002, 1981

A:Title: Adult chicken alpha-globin genes, alpha(A) and alpha(D): no anemic shock alpha-globin gene  
A:Reference number: A93892; MUID:82082384  
A:Accession: B93892  
A:Molecule type: mRNA  
A:Residues: 1-106, 'C', 108, 'RL', 111-141 <DOD>  
R:Takei, H.; Ota, Y.; Wu, K.C.; Kiyohara, T.; Matsuda, G.  
J. Biochem. 77, 1345-1347, 1975  
A:Title: Amino acid sequence of the alpha chain of chicken AI hemoglobin.  
A:Reference number: A91937; MUID:76189993  
A:Accession: A91937  
A:Molecule type: protein  
A:Residues: 1-21, 'Q', 23-37, 'E', 39-52, 'NE', 55-106, 'Q', 108-141 <TAK>

A:Note: this alpha chain is from the adult minor tetrameric component, which has been called alpha-1.  
R:Chapman, B.S.; Hood, L.E.; Tobin, A.J.  
J. Biol. Chem. 257, 651-658, 1982  
A:Title: Minor early embryonic chick hemoglobin M. Amino acid sequences of the epsilon and delta chains.  
A:Reference number: A92377; MUID:82098109  
A:Accession: B92377  
A:Molecule type: protein  
A:Residues: 1-63; 94-106, 'C', 108-110, 'A', 112-122 <CHA>  
A:Experimental source: embryo  
A:Note: this chain was isolated from HbM, the least abundant of the four early chick hemoglobins.

R:Lewis, W.; Lee, J.D.; Dodgson, J.B.  
Nucleic Acids Res. 19, 5321-5329, 1991  
A:Title: Adult chicken alpha-globin gene expression in transfected QT6 quail cells: evidence for alternative splicing  
A:Reference number: S18672; MUID:92020223  
A:Accession: S18672  
A:Status: preliminary; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-15, 'R', 17-141 <LEW>  
A:Cross-references: EMBL:X59989; NID:g63012; PIDN:CAA42605.1; PID:g63013  
C:Genetics:  
A:Gene: alpha-D  
A:Introns: 31/2; 99/3  
C:Superfamily: globin; globin homology  
C:Keywords: blood; chromoprotein; erythrocyte; heme; iron; metalloprotein; oxygen carrier  
F:2-141/Domain: globin homology <GLB>  
F:58/Binding site: oxygen (His) (distal axial ligand) #status predicted  
F:87/Binding site: heme iron (His) (proximal axial ligand) #status predicted

Query Match 2.0%; Score 7; DB 1; Length 141;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 27 VLGALGN 33  
|||||||

Db 62 VLGALGN 68

## RESULT 13

S56103 hemoglobin alpha-D chain - turkey  
C:Species: Meleagris gallopavo (common turkey)  
C:Date: 10-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 04-Mar-2000  
C:Accession: S56103  
R:Equchi, Y.; Ikehara, T.; Kayo, S.; Eguchi, T.; Takei, H.  
Biol. Chem. Hoppe-Seyler 376, 437-440, 1995  
A:Title: Amino acid sequence of alpha- and beta-polypeptide chains of turkey (Meleagris gallopavo) hemoglobin  
A:Reference number: S56102; MUID:96066279  
A:Accession: S56103  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-141 <EGU>  
C:Superfamily: globin; globin homology  
C:Keywords: chromoprotein; heme; iron; metalloprotein; oxygen carrier  
F:2-141/Domain: globin homology <GLB>  
F:58/Binding site: oxygen (His) (distal axial ligand) #status predicted  
F:87/Binding site: heme iron (His) (proximal axial ligand) #status predicted

Query Match 2.0%; Score 7; DB 2; Length 141;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 27 VLGALGN 33  
|||||||

Db 62 VLGALGN 68

## RESULT 14

G64555 hypothetical protein HP0287 - Helicobacter pylori (strain 26695)  
C:Species: Helicobacter pylori  
C:Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 08-Oct-1999  
C:Accession: G64555  
R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKee, P.; Selen, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, son, J.D.; Kelley, J.M.  
Nature 388, 559-547, 1997  
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, A.; Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.  
A:Reference number: A64520; MUID:97394467  
A:Accession: G64555  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-172 <TOM>



A;Cross-references: GB:AE000547; GB:AE000511; NID:g2313377; PIDN:RAD07359.1; PID:g231338  
AG3474

Query Match 2.0%; Score 7; DB 2; Length 172;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 LEFFMPL 189  
|||||||

Db 150 LEFFMPL 156

RESULT 15  
A71953  
hypothetical protein jhp0272 - Helicobacter pylori (strain J99)  
C:Species: Helicobacter pylori  
A:Variety: strain J99  
C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 08-Oct-1999  
C:Accession: A71953  
R;Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;  
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;  
Nature 397, 176-180, 1999  
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path  
A:Reference number: A71800; MUID:99120557  
A:Accession: A71953  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-172 <ARN>  
A:Cross-references: GB:AE001464; GB:AE001439; NID:g4154789; PIDN:RAD05853.1; PID:g415479  
A:Experimental source: strain J99  
C:Genetics:  
A:Gene: jhp0272

Query Match 2.0%; Score 7; DB 2; Length 172;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 LEFFMPL 189  
|||||||

Db 150 LEFFMPL 156

RESULT 16  
F75353  
hypothetical protein - Deinococcus radiodurans (strain RL)  
C:Species: Deinococcus radiodurans  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 26-May-2000  
C:Accession: F75353  
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans RL.  
A:Reference number: A75250; MUID:20036896  
A:Accession: F75353  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-206 <WHI>  
A:Cross-references: GB:AE002021; GB:AE000513; NID:g6459573; PIDN:AAF11349.1; PID:g645957  
A:Experimental source: strain RL  
C:Genetics:  
A:Gene: DR1792  
A:Map position: 1  
C:Superfamily: Deinococcus radiodurans hypothetical protein DR1792

Query Match 2.0%; Score 7; DB 2; Length 206;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 104 VELTVVA 110  
|||||||

Db 163 VELTVVA 169

RESULT 17  
AG3474  
carbamoyl-phosphate synthase large chain (EC 6.3.5.5) [imported] - Brucella melitensis  
C:Species: Brucella melitensis  
C:Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 01-Feb-2002  
C:Accession: AG3474  
R;DelVecchio, V.G.; Kapratral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanov  
; Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Let  
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melit  
A:Reference number: AD3252; PMID:11756688  
A:Accession: AG3474  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-237 <KUR>  
A:Cross-references: GB:AE008917; PIDN:AAL52962.1; PID:g17983813; GSPDB:GN00190  
A:Experimental source: strain 16M  
C:Genetics:  
A:Gene: BME11781  
A:Map position: I  
C:Keywords: ligase

Query Match 2.0%; Score 7; DB 2; Length 237;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 LGNGVAL 37  
|||||||

Db 137 LGNGVAL 143

RESULT 18  
A26599  
clathrin light chain LCA, brain - bovine  
N:Contains: clathrin light chain LCA, lymphocyte  
C:Species: Bos primigenius taurus (cattle)  
C:Date: 31-Mar-1988 #sequence\_revision 31-Mar-1988 #text\_change 13-Aug-1999  
C:Accession: A26599; B26599  
R;Jackson, A.P.; Seow, H.F.; Holmes, N.; Drickamer, K.; Parham, P.  
Nature 326, 154-159, 1987  
A:Title: Clathrin light chains contain brain-specific insertion sequences and a regio  
A:Reference number: A93391; MUID:87144634  
A:Accession: A26599  
A:Molecule type: mRNA  
A:Residues: 1-243 <JAC>  
A:Cross-references: GB:X04849; NID:g507; PIDN:CAA28540.1; PID:g508  
A:Experimental source: brain  
A:Accession: B26599  
A:Molecule type: mRNA  
A:Residues: 1-13, 'H', 15-157, 188-243 <JA2>  
A:Cross-references: GB:X04851; NID:g509; PIDN:CAA28542.1; PID:g510  
A:Experimental source: lymphocyte  
C:Superfamily: clathrin light chain  
C:Keywords: alternative splicing  
F;1-243/Product: clathrin light chain LCA, lymphocyte #status predicted <MAT>  
F;1-157, 188-243/Product: clathrin light chain LCA, lymphocyte #status predicted <MA2>

Query Match 2.0%; Score 7; DB 2; Length 243;  
Best Local Similarity 100.0%; Pred. No. 45;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 ALNGNVA 36  
|||||||

Db 15 ALNGNVA 21

RESULT 19  
A31775  
clathrin light chain A, brain - human  
N:Contains: clathrin light chain A, lymphocyte  
C:Species: Homo sapiens (man)  
C:Date: 21-May-1990 #sequence\_revision 21-May-1990 #text\_change 13-Aug-1999  
C:Accession: A31775; A30752; A30753

R:Jackson, A.P.; Parham, P.  
 J. Biol. Chem. 263, 16688-16695, 1988  
 A:Title: Structure of human clathrin light chains. Conservation of light chain polymorph  
 A:Reference number: A92682; MUID:89034155  
 A:Accession: A31775  
 A:Molecule type: mRNA  
 A:Cross-references: GB:M20471; NID:g179396; PIDN:AAA51817.1; PID:g179397; GB:J04174  
 A:Residues: 1-248 <JAC>  
 A:Accession: A30752  
 A:Status: translation not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-162,193-248 <JA2>  
 A:Cross-references: GB:M20472; NID:g187054; PIDN:AAA59505.1; PID:g3071118; GB:J04174  
 C:Genetics:  
 A:Gene: GDB:CLTA  
 A:Cross-references: GDB:128049; OMIM:118960  
 A:Map position: 12q23-12q24  
 C:Superfamily: clathrin light chain  
 C:Keywords: alternative splicing  
 F:1-248/Product: clathrin light chain A, brain #status predicted <MAT>  
 F:1-162,193-248/Product: clathrin light chain A, lymphocyte #status predicted <MA2>

Query Match 2.08; Score 7; DB 2; Length 248;  
 Best Local Similarity 100.0%; Pred. No. 46;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 30 ALGNQVA 36  
 |||||  
 Db 18 ALGNQVA 24

RESULT 20  
 AH0360  
 Probable dimethyl sulfoxide reductase chain C protein dmsc [imported] - Yersinia pestis  
 C:Species: Yersinia pestis  
 C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Nov-2001  
 C:Accession: AH0360  
 R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.  
 deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;  
 ll, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall,  
 Nature 413, 523-527, 2001  
 A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
 A:Reference number: AB0001; MUID:21470413; PMID:11586360  
 A:Accession: AH0360  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-258 <KUR>  
 A:Cross-references: GB:AL590842; PIDN:CAC92211.1; PID:g15980923; GSPDB:GN00175  
 C:Genetics:  
 A:Gene: dmsC  
 C:Superfamily: probable dimethylsulfoxide reductase chain C

Query Match 2.08; Score 7; DB 2; Length 258;  
 Best Local Similarity 100.0%; Pred. No. 48;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 90 VGLFTLA 96  
 |||||  
 Db 17 VGLFTLA 23

RESULT 21  
 T10530  
 Thymidine deiodinase (EC 3.8.1.4) type II - bullfrog  
 N:Alternate names: diiodothyronine 5'-deiodinase; iodothyronine outer ring monodeiodinase  
 C:Species: Rana catesbeiana (bullfrog)  
 C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 21-Jul-2000  
 C:Accession: T10530  
 R:Davey, J.C.; Becker, K.B.; Schneider, M.J.; St Germain, D.L.; Galton, V.A.  
 J. Biol. Chem. 270, 26786-26789, 1995  
 A:Title: Cloning of a cDNA for the type II iodothyronine deiodinase.  
 A:Reference number: A57768; MUID:96070767  
 A:Accession: T10530

A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-264 <DAV>  
 A:Cross-references: EMBL:LA2815; NID:g1066844; PIDN:AAC42231.1; PID:g1066845  
 C:Function:  
 A:Description: catalyzes the reduction of L-thyroxine by a proton donor to 3',3'',5'-  
 C:Superfamily: vertebrate thyroxine deiodinase I  
 C:Keywords: oxidoreductase; selenocysteine

Query Match 2.08; Score 7; DB 2; Length 264;  
 Best Local Similarity 100.0%; Pred. No. 49;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 58 AVADFLL 64  
 |||||  
 Db 143 AVADFLL 149

RESULT 22  
 T33515  
 hypothetical protein F59H5.3 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 21-Jul-2000  
 C:Accession: T33515  
 R:Jones, K.; Graves, T.; Antoniou, B.  
 submitted to the EMBL Data Library, October 1998  
 A:Description: The sequence of C. elegans cosmid F59H5.  
 A:Reference number: Z21362  
 A:Accession: T33515  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-272 <JON>  
 A:Cross-references: EMBL:AF098991; PIDN:AAC67450.1; GSPDB:GN00020; CESP:F59H5.3  
 A:Experimental source: strain Bristol N2; clone F59H5  
 C:Genetics:  
 A:Gene: CESP:F59H5.3  
 A:Map position: 2  
 A:Introns: 38/2; 168/3; 182/1; 232/3  
 C:Superfamily: Caenorhabditis elegans hypothetical protein F52C6.7; POZ domain homolo

Query Match 2.08; Score 7; DB 2; Length 272;  
 Best Local Similarity 100.0%; Pred. No. 50;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 189 LGIILFC 195  
 |||||  
 Db 49 LGIILFC 55

RESULT 23  
 C82547  
 hypothetical protein XF2517 [imported] - Xylella fastidiosa (strain 9a5c)  
 C:Species: Xylella fastidiosa  
 C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
 C:Accession: C82547  
 R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq  
 Nature 406, 151-157, 2000  
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
 A:Reference number: A82515; MUID:20365717  
 A:Note: for a complete list of authors see reference number A59328 below  
 A:Accession: C82547  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-281 <STM>  
 A:Cross-references: GB:AE004059; GB:AE003849; NID:99107718; PIDN:AAF85315.1; GSPDB:GN  
 A:Experimental source: strain 9a5c  
 R:Simpson, A.J.G.; Reinach, F.C.; Artuda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.  
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carr  
 as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.  
 submitted to GenBank, June 2000  
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr  
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La  
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins

A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;  
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.  
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak  
 A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
 M.; Tshuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
 A;Reference number: A59328  
 C;Contents: annotation  
 C;Genetics:  
 A;Gene: XP2517

Query Match 2.0%; Score 7; DB 2; Length 281;  
 Best Local Similarity 100.0%; Pred. No. 51;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 104 VLTQVA 110  
 |||||  
 DB 186 VLTQVA 192

RESULT 24  
 E97054  
 sporulation protein IVPB related protein, probable metalloproteinase [imported] - Clostr  
 C;Species: Clostridium acetobutylicum  
 C;Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
 C;Accession: E97054  
 R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
 J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
 J. Bacteriol. 183, 4823-4838, 2001  
 A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cl  
 A;Reference number: A96900; MUID:21359325; PMID:21359325  
 A;Accession: E97054  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-284 <KUR>  
 A;Cross-references: GB:AE001437; PIDN:AAK79224.1; PID:g15024179; GSPDB:GN00168  
 A;Experimental source: Clostridium acetobutylicum ATCC824  
 C;Genetics:  
 A;Gene: CAC1253

Query Match 2.0%; Score 7; DB 2; Length 284;  
 Best Local Similarity 100.0%; Pred. No. 52;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 192 ILFCSEK 198  
 |||||  
 DB 183 ILFCSEK 189

RESULT 25  
 LRRTAI  
 clathrin light chain A1 - rat (fragment)  
 N;Alternate names: LCAL  
 C;Species: Rattus norvegicus (Norway rat)  
 C;Date: 31-Mar-1988 #sequence\_revision 31-Mar-1988 #text\_change 10-Dec-1999  
 R;Kirchhausen, T.; Scarmato, P.; Harrison, S.C.; Monroe, J.J.; Chow, E.P.; Mattaliano, R  
 Science 236, 320-324, 1987  
 A;Title: Clathrin light chains LCA and LCB are similar, polymorphic, and share repeated  
 A;Reference number: A94292; MUID:87178007  
 A;Accession: A25994  
 A;Molecule type: mRNA  
 A;Residues: 1-286 <KIR>  
 C;Comment: Clathrin, the major protein component of coated pits and vesicles, is a three  
 r its carboxyl end. The heavy chains are also held together by noncovalent interactions.  
 C;Superfamily: clathrin light chain  
 C;Keywords: alternative splicing; coated pits; coiled coil; endocytosis  
 F:138-238/Domain: coiled coil <HEP>

Query Match 2.0%; Score 7; DB 1; Length 286;  
 Best Local Similarity 100.0%; Pred. No. 52;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 ALNGVA 36  
 |||||  
 DB 56 ALNGVA 62

RESULT 26  
 E83454  
 cell division protein ZlpA PA1528 [imported] - Pseudomonas aeruginosa (strain PA01)  
 C;Species: Pseudomonas aeruginosa  
 C;Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C;Accession: E83454  
 R;Stover, C.K.; Phan, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;  
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L  
 .; Lozy, S.; Olson, M.V.  
 Nature 405, 959-964, 2000  
 A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa  
 A;Reference number: A82950; MUID:20437337  
 A;Accession: E83454  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-289 <STO>  
 A;Cross-references: GB:AE004581; GB:AE004091; NID:g9947482; PIDN:AAG04917.1; GSPDB:GN  
 A;Experimental source: strain PA01  
 C;Genetics:  
 A;Gene: zi; PA1528

Query Match 2.0%; Score 7; DB 2; Length 289;  
 Best Local Similarity 100.0%; Pred. No. 53;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 LPSVSAR 240  
 |||||  
 DB 75 LPSVSAR 81

RESULT 27  
 E71630  
 hypothetical protein RP706 - Rickettsia prowazekii  
 C;Species: Rickettsia prowazekii  
 C;Date: 21-Nov-1998 #sequence\_revision 21-Nov-1998 #text\_change 03-Nov-2000  
 C;Accession: E71630  
 R;Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichert-Ponten, T.; Alsmark  
 Nature 396, 133-140, 1998  
 A;Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.  
 A;Reference number: A71630; MUID:99039499  
 A;Accession: E71630  
 A;Status: preliminary; nucléic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-314 <AND>  
 A;Cross-references: GB:AJ235273; GB:AJ235269; NID:g3861237; PIDN:CAA15141.1; PID:e134  
 A;Experimental source: strain Madrid E  
 C;Genetics:  
 A;Gene: RP706

Query Match 2.0%; Score 7; DB 2; Length 314;  
 Best Local Similarity 100.0%; Pred. No. 56;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 188 PLGIILF 194  
 |||||  
 DB 32 PLGIILF 38

RESULT 28  
 AB3607  
 3-oxoacyl-[acyl-carrier-protein] synthase III (EC 2.3.1.41) [Imported] - Brucella mel  
 C;Species: Brucella melitensis  
 C;Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 15-Feb-2002  
 C;Accession: AB3607  
 R;Delvecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanov  
 .; Mazur, M.; Goldsman, E.; Selkov, P.H.; Hagius, S.; O'Callaghan, D.; Let  
 Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
 A;Title: The genome sequence of the facultative intracellular pathogen Brucella melit

A:Reference number: AD3252; PMID:11756688  
A:Accession: AB3607  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-326 <KUR>  
A:Cross-references: GB:AF008918; PIDN:AAL54021.1; PID:g17984973; GSPDB:GN00191  
A:Experimental source: strain 16M  
C:Genetics:  
A:Gene: BMEI10779  
A:Map position: II  
C:Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase III  
C:Keywords: acyltransferase

Query Match 2.0%; Score 7; DB 2; Length 326;  
Best Local Similarity 100.0%; Pred. No. 58;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 327 ANSFQSQ 333  
|||||||

Db 174 ANSFQSQ 180

## RESULT 29

AI2023  
proteohlorophyllide oxido-reductase [imported] - Anabaena sp. (strain PCC 7120)

C:Species: Anabaena sp.

A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 01-Feb-2002

C:Accession: AI2023

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriiguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S. DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena sp. (strain PCC 7120)

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AI2023

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-329 <KUR>

A:Cross-references: GB:BA000019; PIDN:BA073442.1; PID:g17130833; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: por

C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match 2.0%; Score 7; DB 2; Length 329;  
Best Local Similarity 100.0%; Pred. No. 59;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 LVILGTV 148

|||||||

Db 140 LVILGTV 146

## RESULT 30

I65989

G protein-coupled receptor 5 - human

C:Species: Homo sapiens (man)

C:Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 21-Jul-2000

C:Accession: I65989

R:Heiber, M.; Docherty, J.M.; Shah, G.; Nguyen, T.; Cheng, R.; Heng, H.H.Q.; Marchese, A. DNA Cell Biol. 14, 25-35, 1995

A:Title: Isolation of three novel human genes encoding G protein-coupled receptors.

A:Reference number: I53033; MUID:95134353

A:Accession: I65989

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-333 <RES>

A:Cross-references: GB:L36149; NID:g598154; PIDN:AAA62837.1; PID:g598155

C:Genetics:

A:Gene: GDB:GPR5

A:Cross-references: GDB:371712; OMIM:600552

A:Map position: 3p21.3-3p21.1

C:Superfamily: vertebrate rhodopsin

## Query Match

Best Local Similarity 100.0%; Score 7; DB 2; Length 333;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 189 LGIILFC 195

|||||||

Db 198 LGIILFC 204

## RESULT 31

G64029

hypothetical protein HI1426 - Haemophilus influenzae (strain Rd KW20)

C:Species: Haemophilus influenzae

C:Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 20-Sep-1999

C:Accession: G64029

R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.H.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M. Science 269, 496-512, 1995

A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A. Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.

A:Reference number: A64000; MUID:95350630

A:Accession: G64029

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-340 <TIGR>

A:Cross-references: GB:U32821; GB:L42023; NID:g1574248; PIDN:AAC23063.1; PID:g1574260

C:Genetics:

A:Start codon: GTG

C:Superfamily: ydaa protein

## Query Match

Best Local Similarity 100.0%; Score 7; DB 2; Length 340;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 LVILGTV 148

|||||||

Db 299 LVILGTV 305

## RESULT 32

A42009

N-formyl peptide receptor - human

C:Alternate names: FMLP receptor

C:Species: Homo sapiens (man)

C:Date: 30-Sep-1993 #sequence\_revision 14-Jul-1995 #text\_change 24-Nov-1999

C:Accession: JC2014; A36309; A35495; A42009; I52414

R:Murphy, P.M.; Tiffany, H.L.; Mcdermott, D.; Ahuja, S.K. Gene 133, 285-290, 1993

A:Title: Sequence and organization of the human N-formyl peptide receptor-encoding gene

A:Reference number: JC2014; MUID:94040825

A:Accession: JC2014

A:Molecule type: mRNA

A:Residues: 1-350 <MUR>

A:Cross-references: GB:L10820; NID:g182739; PIDN:AAA16863.1; PID:g182740

A:Note: 192-Asn was also found

R:Boulay, F.; Tardif, M.; Brouchon, L.; Vignais, P. Biochemistry 29, 11123-11133, 1990

A:Title: The human N-formylpeptide receptor. Characterization of two cDNA isolates an

A:Reference number: A36309; MUID:91105045

A:Accession: A36309

A:Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-191, 'N', 193-345, 'E', 347-350 <BOU1>

A:Cross-references: GB:M60627; GB:M33538; NID:g182664; PIDN:AAA35847.1; PID:g182665

R:Boulay, F.; Tardif, M.; Brouchon, L.; Vignais, P. Biochem. Biophys. Res. Commun. 168, 1103-1109, 1990

A:Title: Synthesis and use of a novel N-formyl peptide derivative to isolate a human

A:Reference number: A35495; MUID:90267449

A:Accession: A35495

A:Molecule type: mRNA

A:Residues: 1-100, 'L', 102-191, 'N', 193-350 <BOU2>

A:Cross-references: GB:M37128; NID:g189183  
 A:Note: the sequence in GenBank entry HUNNFRP, release 112.0, (PIDN:AAA36362.1) has the  
 R:BaO, L.; Gerard, N.P.; Eddy Jr., R.L.; Shows, T.B.; Gerard, C.  
 Genomics 13, 437-440, 1992  
 A:Title: Mapping of genes for the human C5a receptor (C5AR), human FMLP receptor (FPR),  
 A:Reference number: A42009; MUID:92307681  
 A:Accession: A42009  
 A:Status: nucleic acid sequence not shown; not compared with conceptual translation  
 A:Molecule type: DNA  
 A:Residues: 1,'G',3-28,30-100,'L',102-104,106-112,'FLTA',115-176,178-182,184-191,'N',194  
 B:Perez, H.D.; Holmes, R.; Kelly, E.; McClary, J.; Chou, Q.; Andrews, W.H.  
 Biochemistry 31, 11595-11599, 1992  
 A:Title: Cloning of the gene coding for a human receptor for formyl peptides. Characteri  
 A:Reference number: I52414; MUID:93075765  
 A:Accession: I52414  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-5 <PER>  
 A:Cross-references: GB:S49810; NID:g260832; PIDN:AAD14906.1; PID:g4262758  
 C:Genetics:  
 A:Gene: GDB:FPRI  
 A:Cross-references: GDB:127999; OMIM:136537  
 A:Map position: 19q13.4-19q13.4  
 A:Introns: #status absent  
 A:Note: entire coding region is found in exon 3; alternatively spliced mRNA transcripts  
 C:Superfamily: chemotaxis; G protein-coupled receptor; glycoprotein; transmembrane protein  
 F:27-53/Domain: transmembrane #status predicted <TM1>  
 F:59-83/Domain: transmembrane #status predicted <TM2>  
 F:100-121/Domain: transmembrane #status predicted <TM3>  
 F:145-169/Domain: transmembrane #status predicted <TM4>  
 F:206-225/Domain: transmembrane #status predicted <TM5>  
 F:242-266/Domain: transmembrane #status predicted <TM6>  
 F:279-304/Domain: transmembrane #status predicted <TM7>  
 F:4,10/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 F:98-176/Disulfide bonds: #status predicted

Query Match 2.0%; Score 7; DB 2; Length 350;  
 Best Local Similarity 100.0%; Pred. No. 62;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 56 NLAVADF 62  
 Db 66 NLVAADF 72  
 |||||

RESULT 33  
 I49339  
 macrophage inflammatory protein-1 alpha receptor - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 13-Aug-1999  
 C:Accession: I49339  
 R:Gao, J.L.; Murphy, P.M.  
 J. Biol. Chem. 270, 17494-17501, 1995  
 A:Title: Cloning and differential tissue-specific expression of three mouse beta chemok  
 A:Reference number: I49339; MUID:95340546  
 A:Accession: I49339  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-355 <RES>  
 A:Cross-references: EMBL:U28404; NID:g881547; PIDN:AAA89153.1; PID:g881548  
 C:Superfamily: vertebrate rhodopsin

Query Match 2.0%; Score 7; DB 2; Length 355;  
 Best Local Similarity 100.0%; Pred. No. 63;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 53 YLFNLAV 59  
 Db 72 YLFNLAV 78  
 |||||

RESULT 34

B71309  
 hypothetical protein TP0572 - syphilis spirochete  
 C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)  
 C:Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 05-Nov-1999  
 C:Accession: B71309  
 R:Fraser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson, R.; G  
 rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; M  
 they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.  
 Science 281, 375-388, 1998  
 A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.  
 A:Reference number: A71250; MUID:98332770  
 A:Accession: B71309  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-360 <COL>  
 A:Cross-references: GB:AE001232; GB:AE000520; NID:g3322856; PIDN:AAC65551.1; PID:g332  
 C:Genetics:  
 A:Gene: TP0572  
 C:Superfamily: syphilis spirochete hypothetical protein TP0572

Query Match 2.0%; Score 7; DB 2; Length 360;  
 Best Local Similarity 100.0%; Pred. No. 63;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 24 VAFVLGA 30  
 Db 239 VAFVLGA 245  
 |||||

RESULT 35  
 A49092  
 angiotensin II receptor type 2 - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 17-Nov-1995 #sequence\_revision 17-Nov-1995 #text\_change 20-Jun-2000  
 C:Accession: A49092; A49093  
 R:Mukoyama, M.; Nakajima, M.; Horiuchi, M.; Sasamura, H.; Pratt, R.E.; Dzau, V.J.  
 J. Biol. Chem. 268, 24539-24542, 1993  
 A:Title: Expression cloning of type 2 angiotensin II receptor reveals a unique class  
 A:Reference number: A49092; MUID:94043302  
 A:Accession: A49092  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-363 <MUK>  
 A:Cross-references: GB:U01908; NID:g430987; PIDN:AAC52126.1; PID:g430988  
 R:Kambayashi, Y.; Bardhan, S.; Takahashi, K.; Tsuzuki, S.; Inui, H.; Hamakubo, T.; In  
 J. Biol. Chem. 268, 24543-24546, 1993  
 A:Title: Molecular cloning of a novel angiotensin II receptor isoform involved in pho  
 A:Reference number: A49093; MUID:94043303  
 A:Accession: A49093  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-363 <KAM>  
 A:Cross-references: GB:DL6840; NID:g432365; PIDN:BAA04116.1; PID:g432366  
 C:Superfamily: vertebrate rhodopsin

Query Match 2.0%; Score 7; DB 2; Length 363;  
 Best Local Similarity 100.0%; Pred. No. 64;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 55 FNLAVAD 61  
 Db 84 FNLAVAD 90  
 |||||

RESULT 36  
 JC2543  
 angiotensin II receptor type 2 - human  
 C:Species: Homo sapiens (man)  
 C:Date: 13-Jun-1995 #sequence\_revision 14-Jul-1995 #text\_change 24-Nov-1999  
 C:Accession: JC2543; JC2228; S55523; I38943; JC2435  
 R:Kolke, G.; Horiuchi, M.; Yamada, T.; Szpirer, C.; Jacob, H.J.; Dzau, V.J.  
 Biochem. Biophys. Res. Commun. 203, 1842-1850, 1994

A:Title: Human type 2 angiotensin II receptor gene: Cloned, mapped to the X chromosome,  
 A:Reference number: JC2543; MUID:95032069  
 A:Accession: JC2543  
 A:Molecule type: DNA  
 A:Residues: 1-363 <KO>  
 A:Cross-references: EMBL:U10273; NID:9607811; PIDN:AAA61794.1; PID:9607812  
 A:Experimental source: leukocytes  
 R:Tsuchi, S.; Ichiki, T.; Nakakubo, H.; Kitami, Y.; Guo, D.F.; Shirai, H.; Inagami, T.  
 Biochem. Biophys. Res. Commun. 200, 1449-1454, 1994  
 A:Title: Molecular cloning and expression of the gene encoding human angiotensin II type  
 A:Reference number: JC2228; MUID:94242007  
 A:Accession: JC2228  
 A:Molecule type: DNA  
 A:Residues: 1-363 <TSU>  
 A:Cross-references: GB:U15592; NID:9558882; PIDN:AAA50762.1; PID:9558883  
 A:Note: the sequence in GenBank entry HSU15592, release 111.0, (PID:9558883) has the cod  
 R:Warnecke, C.H.; Holmeister, J.; Regitz-Zagrosek, V.; Fleck, E.  
 submitted to the EMBL Data Library, June 1995  
 A:Description: Cloning and characterisation of the human angiotensin II receptor type 2  
 A:Reference number: S5523  
 A:Accession: S5523  
 A:Molecule type: mRNA  
 A:Residues: 1-16 <WAR>  
 A:Cross-references: EMBL:X87723; NID:9860958; PIDN:CAA61022.1; PID:9860959  
 R:Martin, M.M.; Elton, T.S.  
 Biochem. Biophys. Res. Commun. 209, 554-562, 1995  
 A:Title: The sequence and genomic organization of the human type 2 angiotensin II recept  
 A:Reference number: I38943; MUID:95251653  
 A:Accession: I38943  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-363 <MAR1>  
 A:Cross-references: EMBL:U20860; NID:9747969; PIDN:AAA85851.1; PID:9747970  
 R:Martin, M.M.; Su, B.; Elton, T.S.  
 Biochem. Biophys. Res. Commun. 205, 645-651, 1994  
 A:Title: Molecular cloning of the human angiotensin II type 2 receptor cDNA.  
 A:Reference number: JC2435; MUID:95091796  
 A:Accession: JC2435  
 A:Molecule type: mRNA  
 A:Residues: 1-267, 'WC', 270-363 <MAR2>  
 A:Cross-references: GB:U16957; NID:9595934; PIDN:AAA67753.1; PID:9595935  
 A:Experimental source: lung  
 A:Note: the authors translated the codon TGG for residue 268 as Cys and TGC for residue  
 C:Genetics:  
 A:Gene: GDB:AGTR2  
 A:Cross-references: GDB:I34188; OMIM:300034  
 A:Map position: Xq22-Xq23  
 C:Superfamily: vertebrate rhodopsin  
 C:Keywords: glycoprotein; phosphoprotein; receptor; transmembrane protein  
 F:46-69/Domain: transmembrane #status predicted <TM1>  
 F:81-104/Domain: transmembrane #status predicted <TM2>  
 F:119-140/Domain: transmembrane #status predicted <TM3>  
 F:158-181/Domain: transmembrane #status predicted <TM4>  
 F:209-234/Domain: transmembrane #status predicted <TM5>  
 F:258-281/Domain: transmembrane #status predicted <TM6>  
 F:292-308/Domain: transmembrane #status predicted <TM7>  
 F:4,13,24,29,34/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 F:79,346,353/Binding site: phosphate (Ser) (covalent) (by CAMP- and cGMP-dependent kinase  
 F:152,348,354/Binding site: phosphate (Ser) (covalent) (by protein kinase C) #status pre  
 F:354/Binding site: phosphate (Ser) (covalent) (by protein kinase C) #status predicted

Query Match 2.0%; Score 7; DB 2; Length 363;  
 Best Local Similarity 100.0%; Pred. No. 64;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 FNLAVID 61  
 Db 84 FNLAVID 90  
 |||||

RESULT 37  
 T41929  
 polymerase processivity factor - human herpesvirus 7 (strain JI)

C:Species: human herpesvirus 7  
 A:Variety: strain JI  
 A:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jan-2000  
 C:Accession: T41929  
 R:Nicholas, J.  
 submitted to the EMBL Data Library, December 1995  
 A:Description: Determination and analysis of the complete nucleotide sequence of huma  
 A:Reference number: Z22022  
 A:Accession: T41929  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-364 <NIC>  
 A:Cross-references: EMBL:U43400; PIDN:AAC54689.1  
 A:Experimental source: strain JI  
 C:Genetics:  
 A:Note: U27  
 C:Superfamily: human herpesvirus 6 P41 protein

Query Match 2.0%; Score 7; DB 2; Length 364;  
 Best Local Similarity 100.0%; Pred. No. 64;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 42 FHMKTWK 48  
 Db 28 FHMKTWK 34  
 |||||

RESULT 38  
 B83313  
 probable type II secretion system protein PA2676 [Imported] - Pseudomonas aeruginosa  
 C:Species: Pseudomonas aeruginosa  
 C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C:Accession: B83313  
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;  
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L  
 ; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa  
 A:Reference number: A82950; MUID:20437337  
 A:Accession: B83313  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-395 <STO>  
 A:Cross-references: GB:AE004695; GB:AE004091; NID:99948730; PIDN:AAG06064.1; GSPDB:GN  
 A:Experimental source: strain PA01  
 C:Genetics:  
 A:Gene: PA2676

Query Match 2.0%; Score 7; DB 2; Length 395;  
 Best Local Similarity 100.0%; Pred. No. 69;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 212 QARMKKA 218  
 Db 354 QARMKKA 360  
 |||||

RESULT 39  
 T49810  
 hypothetical protein B11B22.120 [Imported] - Neurospora crassa  
 C:Species: Neurospora crassa  
 C:Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 02-Jun-2000  
 C:Accession: T49810  
 R:Schulte, U.; Alpn, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatu  
 submitted to the Protein Sequence Database, May 2000  
 A:Reference number: Z25022  
 A:Accession: T49810  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-400 <SCH>  
 A:Cross-references: EMBL:AL356834; GSPDB:GN00116; NCSP:B11B22.120  
 A:Experimental source: BAC clone B11B22; strain OR74A  
 C:Genetics:

A:Gene: NCSP:B11B22.120  
A:Map position: 6

Query Match 2.0%; Score 7; DB 2; Length 400;  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 LLIVAFV 27  
|||||||  
Db 138 LLIVAFV 144

RESULT 40

A:Title: hypothetical protein BBA57 - Lyme disease spirochete plasmid A/1p54  
C:Species: Borrelia burgdorferi (Lyme disease spirochete)  
C:Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 08-Oct-1999  
C:Accession: A70214  
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White  
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt  
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.  
Nature 390, 580-586, 1997  
A:Authors: Smith, H.O.; Venter, J.C.  
A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.  
A:Reference number: A70100; MUID:98065943.  
A:Accession: A70214  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-414 <KLE>  
A:Cross-references: GB:AE000790; NID:g2690224; PIDN:AAC66270.1; PID:g2690269; TIGR:BBA57  
A:Experimental source: strain B31  
C:Genetics:  
A:Genome: plasmid

Query Match 2.0%; Score 7; DB 2; Length 414;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 286 PKFYNKL 292  
|||||||  
Db 345 PKFYNKL 351

RESULT 41

AG1804  
sugar transport protein homolog lin2982 [imported] - Listeria innocua (strain Clip11262)  
C:Species: Listeria innocua  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
C:Accession: AG1804  
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihl, H.  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla  
A:Title: Comparative genomics of Listeria species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AG1804  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-422 <GLA>  
A:Cross-references: GB:AL592022; PIDN:CAC98207.1; PID:g16415522; GSPDB:GN00178  
A:Experimental source: strain Clip11262  
C:Genetics:  
A:Gene: lin2982

Query Match 2.0%; Score 7; DB 2; Length 422;  
Best Local Similarity 100.0%; Pred. No. 73;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 LLIVAFV 27  
|||||||  
Db 173 LLIVAFV 179

RESULT 42

AI1430  
sugar transport proteins homolog lmo2850 [imported] - Listeria monocytogenes (strain  
C:Species: Listeria monocytogenes  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
C:Accession: AI1430  
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec  
; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihl,  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.;  
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla  
A:Title: Comparative genomics of Listeria species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AI1430  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-422 <GLA>  
A:Cross-references: GB:NC\_003210; PIDN:CAD01063.1; PID:g16412350; GSPDB:GN00177  
A:Experimental source: strain EGD-e  
C:Genetics:  
A:Gene: lmo2850

Query Match 2.0%; Score 7; DB 2; Length 422;  
Best Local Similarity 100.0%; Pred. No. 73;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 LLIVAFV 27  
|||||||  
Db 173 LLIVAFV 179

RESULT 43

T27955  
hypothetical protein ZK669.4 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 31-Jan-2000  
C:Accession: T27955  
R:Thomas, K.  
submitted to the EMBL Data Library, September 1994  
A:Reference number: Z20446  
A:Accession: T27955  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-448 <WIL>  
A:Cross-references: EMBL:Z37093; PIDN:CAA85465.1; GSPDB:GN00020; CESP:ZK669.4  
A:Experimental source: clone ZK669  
C:Genetics:  
A:Gene: CESP:ZK669.4  
A:Map position: 2

Query Match 2.0%; Score 7; DB 2; Length 448;  
Best Local Similarity 100.0%; Pred. No. 76;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 EGDITSQ 16  
|||||||  
Db 54 EGDITSQ 60

RESULT 44

AH0708  
PTS system, cellobiose-specific IIC component [imported] - Salmonella enterica subsp.  
C:Species: Salmonella enterica subsp. enterica serovar Typhi  
A:Note: this species has also been called Salmonella typhi  
C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 27-Nov-2001  
C:Accession: AH0708  
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church  
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr

, S.; Moule, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A:Authors: Parry, C.; Ouail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov  
A:Reference number: AB0502; PMID:11677608  
A:Accession: AH0708  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-452 <PAR>  
A:Cross-references: GB:AL513382; PIDN:CAD02041.1; PID:gl6502878; GSPDB:GN00176  
C:Genetics:  
A:Gene: celB  
C:Superfamily: phosphotransferase system enzyme II factor II, phosphoenolpyruvate-depend

Query Match 2.0%; Score 7; DB 2; Length 452;  
Best Local Similarity 100.0%; Pred. No. 77;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 LFNLA 60  
|||||  
Db 413 LFNLA 419

RESULT 45  
T24686  
hypothetical protein T08D10.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 04-Mar-2000  
C:Accession: T24686  
R:Lloyd, C.  
submitted to the EMBL Data Library, August 1995  
A:Reference number: Z19923  
A:Accession: T24686  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-473 <WIL>  
A:Cross-references: EMBL:Z50756; PIDN:CAA90638.1; GSPDB:GN00028; CESP:T08D10.3  
A:Experimental source: clone T08D10  
C:Genetics:  
A:Gene: CESP:T08D10.3  
A:Map position: X  
A:Introns: 27/3; 87/3; 161/3; 266/2; 447/1  
C:Superfamily: Caenorhabditis elegans hypothetical protein T08D10.3

Query Match 2.0%; Score 7; DB 2; Length 473;  
Best Local Similarity 100.0%; Pred. No. 80;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 203 LRRRQQL 209  
|||||  
Db 204 LRRRQQL 210

Search completed: October 30, 2002, 18:27:34  
Job time : 22 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2002, 18:26:03 ; Search time 13 Seconds  
(without alignments)  
650.097 Million cell updates/sec

Title: US-09-886-041-2  
Perfect score: 346  
Sequence: 1 MYNSCCRIEQTISQVMP.....ANSFQSQDQWDPHIVEH 346

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 231628 seqs, 2442594 residues

Word size : 0

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	2.3	350	2	US-08-458-970A-9
2	7	2.0	315	1	US-08-118-270-34
3	7	2.0	315	5	PCT-US93-08528-34
4	7	2.0	350	1	US-07-759-568-3
5	7	2.0	350	4	US-08-430-286A-8
6	7	2.0	353	1	US-08-148-209A-3
7	7	2.0	495	3	US-09-079-415-4
8	7	2.0	549	2	US-08-676-279-59
9	7	2.0	744	3	US-08-462-080B-2
10	7	2.0	744	3	US-08-462-090-2
11	7	2.0	744	3	US-08-463-461-2
12	7	2.0	777	2	US-08-874-678-3
13	7	2.0	777	3	US-08-643-839-3
14	7	2.0	941	4	US-09-179-558-55
15	7	2.0	1298	1	US-08-222-616-33
16	7	2.0	1298	1	US-08-340-011-2
17	7	2.0	1298	3	US-08-901-710-2
18	7	2.0	1298	5	PCT-US95-04228-33
19	7	2.0	1362	2	US-08-874-678-33
20	7	2.0	1362	3	US-08-643-839-33
21	7	2.0	1363	1	US-08-340-011-4
22	7	2.0	1363	2	US-08-874-678-32
23	7	2.0	1363	3	US-08-643-839-32
24	7	2.0	1368	3	US-08-901-710-4
25	7	2.0	1368	2	US-08-874-678-34
26	7	2.0	1368	3	US-08-643-839-34
27	6	1.7	9	3	US-08-491-954-21

28	6	1.7	11	5	PCT-US92-03432-3	Sequence 3, Appl1
29	6	1.7	18	1	US-07-798-223A-2	Sequence 2, Appl1
30	6	1.7	24	6	5187078-2	Patent No. 5187078
31	6	1.7	40	1	US-08-383-751A-3	Sequence 3, Appl1
32	6	1.7	56	2	US-08-637-759B-358	Sequence 358, App
33	6	1.7	56	3	US-08-871-355A-358	Sequence 358, App
34	6	1.7	56	4	US-09-201-945-358	Sequence 34, Appl
35	6	1.7	70	3	US-08-513-974B-34	Sequence 317, App
36	6	1.7	70	3	US-08-513-974B-317	Sequence 14, Appl
37	6	1.7	80	1	US-07-971-160-14	Sequence 14, Appl
38	6	1.7	80	1	US-08-336-241-14	Sequence 14, Appl
39	6	1.7	80	2	US-08-465-273-14	Sequence 14, Appl
40	6	1.7	80	2	US-09-119-024-14	Sequence 14, Appl
41	6	1.7	80	2	US-08-417-226-14	Sequence 14, Appl
42	6	1.7	80	4	US-09-196-131-14	Sequence 14, Appl
43	6	1.7	82	4	US-08-961-083-130	Sequence 130, App
44	6	1.7	88	1	US-08-091-569-21	Sequence 21, Appl
45	6	1.7	88	1	US-08-203-676-21	Sequence 21, Appl

ALIGNMENTS

RESULT 1  
US-08-458-970A-9  
; Sequence 9, Application US/08458970A  
; Patent No. 5861272  
; GENERAL INFORMATION:  
; APPLICANT: LI, ET AL.  
; TITLE OF INVENTION: C5a Receptor  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
; ADDRESSEE: CECCHI, STEWART & OLSTEIN  
; STREET: 6 BECKER FARM ROAD  
; CITY: ROSELAND  
; STATE: NEW JERSEY  
; COUNTRY: USA  
; ZIP: 07068  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 INCH DISKETTE  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WORD PERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/458,970A  
; FILING DATE: June 2, 1995  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US94/09234  
; FILING DATE: 16 AUG 1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MULLINS, J.G.  
; REGISTRATION NUMBER: 33,073  
; REFERENCE/DOCKET NUMBER: 325800-353  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-994-1700  
; TELEFAX: 201-994-1744  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 350 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-458-970A-9

Query Match 2.3%; Score 8; DB 2; Length 350;  
Best Local Similarity 100.0%; Pred. No. 5.5;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 56 NLAVALDF 63  
|||||||

Db 77 NLAADF 84

## RESULT 2

US-08-118-270-34  
; Sequence 34, Application US/08118270  
; Patent No. 5508384  
; GENERAL INFORMATION:  
; APPLICANT: Murphy, Randall B.  
; ATTORNEY/AGENT INFORMATION:  
; TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN  
; TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF  
; NUMBER OF SEQUENCES: 348  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/118,270  
FILING DATE: 09-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/943,236  
FILING DATE: 10-SEP-1992

ATTORNEY/AGENT INFORMATION:  
NAME: Townsend, Kevin G.  
REGISTRATION NUMBER: 34,033  
REFERENCE/DOCKET NUMBER: MURPHY-2A

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633

INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 315 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide

US-08-118-270-34

Query Match 2.0%; Score 7; DB 1; Length 315;  
Best Local Similarity 100.0%; Pred. No. 49;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 56 NLAADF 62  
Db 40 NLAADF 46

## RESULT 3

PCT-US93-08528-34  
; Sequence 34, Application PC/TUS9308528  
; GENERAL INFORMATION:  
; APPLICANT: New York University  
; TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN  
; TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF  
; NUMBER OF SEQUENCES: 348  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/08528  
FILING DATE: 09-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/943,236  
FILING DATE: 10-SEP-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Townsend, Kevin G.  
REGISTRATION NUMBER: 34,033  
REFERENCE/DOCKET NUMBER: MURPHY-2 PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633

INFORMATION FOR SEQ ID NO: 34:

SEQUENCE CHARACTERISTICS:

LENGTH: 315 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

PCT-US93-08528-34

Query Match 2.0%; Score 7; DB 5; Length 315;

Best Local Similarity 100.0%; Pred. No. 49;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 56 NLAADF 62

Db 40 NLAADF 46

## RESULT 4

US-07-759-568-3  
; Sequence 3, Application US/07759568  
; Patent No. 5374506  
; GENERAL INFORMATION:  
; APPLICANT: Murphy, Phillip M.  
; TITLE OF INVENTION: Cloning of cDNA Encoding a Functional  
; TITLE OF INVENTION: Human Interleukin-8 Receptor  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cushman, Darby & Cushman  
; STREET: 1615 L Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036-5601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/759,568

FILING DATE: 19910913

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Scott, Watson T.

REGISTRATION NUMBER: 26581

REFERENCE/DOCKET NUMBER: WTS/5683/91535/WB

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-861-3000

TELEFAX: 202-822-0944

TELEX: 6714627 cush

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 350 amino acids

TYPE: AMINO ACID

; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-07-759-568-3

Query Match 2.0%; Score 7; DB 1; Length 350;  
Best Local Similarity 100.0%; Pred. No. 54;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 NLAADF 62

|||||||

Db 66 NLAADF 72

RESULT 5

US-08-430-286A-8

; Sequence 8, Application US/08430286A

; Patent No. 6225080

; GENERAL INFORMATION:

; APPLICANT: Uhl, George R.

; APPLICANT: Eppler, C. Mark

; APPLICANT: Wang, Jai-Bel

; TITLE OF INVENTION: Mu-Subtype Opioid Receptor

; NUMBER OF SEQUENCES: 14

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Darby & Darby PC

; STREET: 805 Third Avenue

; CITY: New York

; STATE: New York

; COUNTRY: US

; ZIP: 10022

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/430,286A

; FILING DATE: 28-APR-1995

; CLASSIFICATION: 536

; ATTORNEY/AGENT INFORMATION:

; NAME: Robinson, Joseph R.

; REGISTRATION NUMBER: 33,448

; REFERENCE/DOCKET NUMBER: 0646/1A843-US5

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-527-7700

; TELEFAX: 212-753-6237

; TELEX: 236687

; INFORMATION FOR SEQ ID NO: 8:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 350 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; FRAGMENT TYPE: N-terminal

; ORIGINAL SOURCE:

; ORGANISM: Homo sapiens

; IMMEDIATE SOURCE:

; CLONE: F-PEP

US-08-430-286A-8

Query Match 2.0%; Score 7; DB 4; Length 350;  
Best Local Similarity 100.0%; Pred. No. 54;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 NLAADF 62

|||||||

Db 66 NLAADF 72

RESULT 6

US-08-148-209A-3

; Sequence 3, Application US/08148209A

; Patent No. 5556780  
; GENERAL INFORMATION:  
; APPLICANT: Dzau, Victor J  
; APPLICANT: Mukoyama, Masashi  
; TITLE OF INVENTION: TYPE-2 ANGIOTENSIN II RECEPTOR AND GENE  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-4187

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/148,209A

; FILING DATE: 05-NOV-1993

; CLASSIFICATION: 800

; ATTORNEY/AGENT INFORMATION:

; NAME: Rowland, Bertram I

; REGISTRATION NUMBER: 20,015

; REFERENCE/DOCKET NUMBER: A-58491-1/BIR

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 781-1989

; TELEFAX: (415) 398-3249

; TELEX: 910 277299

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 363 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

US-08-148-209A-3

Query Match 2.0%; Score 7; DB 1; Length 363;

Best Local Similarity 100.0%; Pred. No. 56;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 FNLAADF 61

|||||||

Db 84 FNLAADF 90

RESULT 7

US-09-079-415-4

; Sequence 4, Application US/09079415

; Patent No. 6013452

; GENERAL INFORMATION:

; APPLICANT: Christensen, Tove

; APPLICANT: Lehbeck, Jan

; TITLE OF INVENTION: A Fungus Wherein The areaA, pepC and/or

; TITLE OF INVENTION: pepE Genes Have been Inactivated

; NUMBER OF SEQUENCES: 12

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 6013452o No. 6013452disk of No. 6013452th America, Inc.

; STREET: 405 Lexington Avenue

; CITY: New York

; STATE: NY

; COUNTRY: USA

; ZIP: 10174

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/079,415

; FILING DATE: 14-MAY-1998

CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Rozek, Carol  
REGISTRATION NUMBER: 36,993  
REFERENCE/DOCKET NUMBER: 4657,204-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 495 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-079-415-4

Query Match 2.0%; Score 7; DB 3; Length 495;  
Best Local Similarity 100.0%; Pred. No. 74;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 122 HAVNTIS 128  
|||||||  
Db 370 HAVNTIS 376

RESULT 8  
US-08-676-279-59  
Sequence 59, Application US/08676279  
Patent No. 5869247  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: MACROPHAGE NUCLEOTIDE SEQUENCE  
NUMBER OF SEQUENCES: 63  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/676,279  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB95/00095  
APPLICATION NUMBER: GB 9400929.7  
FILING DATE: 19-JAN-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9422021.7  
FILING DATE: 31-OCT-1994  
INFORMATION FOR SEQ ID NO: 59:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 549 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-676-279-59

Query Match 2.0%; Score 7; DB 2; Length 549;  
Best Local Similarity 100.0%; Pred. No. 81;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 133 AGIVCTL 139  
|||||||  
Db 377 AGIVCTL 383

RESULT 9  
US-08-462-080B-2  
Sequence 2, Application US/08462080B  
Patent No. 5997913  
GENERAL INFORMATION:  
APPLICANT: Fowler, Timothy  
APPLICANT: Barnett, Christopher C.  
APPLICANT: Shoemaker, Sharon

TITLE OF INVENTION: Saccharification of Cellulose by Cloning and  
TITLE OF INVENTION: Amplification of the Beta-glucosidase Gene of Trichoderma R  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genencor International, Inc.  
STREET: 925 Page Mill Road  
CITY: Palo Alto  
STATE: Ca  
COUNTRY: U.S.A.  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/462,080B  
FILING DATE: 05-JUN-1995  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/248,586  
FILING DATE: 24-MAY-1994  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/807,028  
FILING DATE: 10-DEC-1991  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/625,140  
FILING DATE: 10-DEC-1990  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Stone, Christopher L.  
REGISTRATION NUMBER: 35,696  
REFERENCE/DOCKET NUMBER: GC78D3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-846-7555  
TELEFAX: 650-845-6504  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 744 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-462-080B-2

Query Match 2.0%; Score 7; DB 2; Length 744;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 225 VAIVFIT 231  
|||||||  
Db 461 VAIVFIT 467

RESULT 10  
US-08-462-090-2  
Sequence 2, Application US/08462090  
Patent No. 602725  
GENERAL INFORMATION:  
APPLICANT: Fowler, Timothy  
APPLICANT: Barnett, Christopher C.  
APPLICANT: Shoemaker, Sharon  
TITLE OF INVENTION: Saccharification of Cellulose by Cloning  
TITLE OF INVENTION: and Amplification of the Beta-glucosidase Gene of  
TITLE OF INVENTION: Trichoderma Reesel  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: George Mason Building, 699 Prince St.  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: U.S.A.

; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/462,090  
; FILING DATE: 05-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/625,140  
; FILING DATE: 10-DEC-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dillahunty, T. Gene  
; REGISTRATION NUMBER: 25,423  
; REFERENCE/DOCKET NUMBER: 010055-056  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-854-7400  
; TELEFAX: 415-854-8275  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 744 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-462-090-2

Query Match 2.0%; Score 7; DB 3; Length 744;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 225 VAIVFIT 231  
Db 461 VAIVFIT 467

RESULT 11  
US-08-463-461-2  
; Sequence 2, Application US/08463461  
; Patent No. 6103464  
; GENERAL INFORMATION:  
; APPLICANT: Fowler, Timothy  
; APPLICANT: Barnett, Christopher C.  
; APPLICANT: Shoemaker, Sharon  
; TITLE OF INVENTION: Saccharification of Cellulose by Cloning  
; TITLE OF INVENTION: and Amplification of the Beta-glucosidase Gene of  
; TITLE OF INVENTION: Trichoderma Reesei  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genencor International, Inc.  
; STREET: 925 Page Mill Road  
; CITY: Palo Alto  
; STATE: California  
; ZIP: 94304-1013  
; COUNTRY: U.S.A.  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/463,461  
; FILING DATE: 05-JUN-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Christopher L. Stone  
; REGISTRATION NUMBER: 35,696  
; REFERENCE/DOCKET NUMBER: GC78D4  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-846-7555  
; TELEFAX: 415-845-6504  
; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 744 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-463-461-2  
Query Match 2.0%; Score 7; DB 3; Length 744;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 225 VAIVFIT 231  
Db 461 VAIVFIT 467

RESULT 12  
US-08-874-678-3  
; Sequence 3, Application US/08874678  
; Patent No. 5952199  
; GENERAL INFORMATION:  
; APPLICANT: Davis-Smyth, Terri L.  
; APPLICANT: Chen, Helen H.  
; APPLICANT: Presta, Leonard  
; APPLICANT: Ferrara, Napoleone  
; TITLE OF INVENTION: NOVEL INHIBITORS OF VASCULAR ENDOTHELIAL GROWTH FACTOR  
; TITLE OF INVENTION: ACTIVITY, THEIR USES AND PROCESSES FOR THEIR PRODUCTION  
; NUMBER OF SEQUENCES: 48  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
; STREET: Four Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/874,678  
; FILING DATE: HEREWITH  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/643,839  
; FILING DATE: 07-MAY-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dreger, Walter H.  
; REGISTRATION NUMBER: 24,190  
; REFERENCE/DOCKET NUMBER: A-63291-1/WHH  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 777 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-08-874-678-3

Query Match 2.0%; Score 7; DB 2; Length 777;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 202 SLRRRQ 208  
Db 473 SLRRRQ 479

```
RESULT 13
US-08-643-839-3
; Sequence 3, Application US/08643839
; Patent No. 6100071
; GENERAL INFORMATION:
; APPLICANT: Davis-Smyth, Terri L.
; APPLICANT: Chen, Helen H.
; APPLICANT: Presta, Leonard
; APPLICANT: Ferrara, Napoleone
; TITLE OF INVENTION: NOVEL INHIBITORS OF VASCULAR ENDOTHELIAL
; TITLE OF INVENTION: GROWTH FACTOR ACTIVITY, THEIR USES AND PROCESSES FOR THEIR
; TITLE OF INVENTION: PRODUCTION
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/643,839
; FILING DATE: 07-MAY-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Walter H.
; REGISTRATION NUMBER: 24,190
; REFERENCE/DOCKET NUMBER: A-63291/WHB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 777 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-643-839-3

Query Match 2.0%; Score 7; DB 3; Length 777;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 202 SURRRQQ 208
DB 473 SURRRQQ 479

RESULT 14
US-09-179-558-55
; Sequence 55, Application US/09179558
; Patent No. 6180612
; GENERAL INFORMATION:
; APPLICANT: Hockensmith, Joel W.
; APPLICANT: Muthuswami, Rohini
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
; TITLE OF INVENTION: TARGETING DNA METABOLIC PROCESSES USING
; TITLE OF INVENTION: AMINOGLYCOSIDE DERIVATIVES
; NUMBER OF SEQUENCES: 66
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
```

```
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/179,558
; FILING DATE: 27-OCT-1998
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. 09/060,470
; FILING DATE: 15-APR-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. 60/063,898
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 9426-005-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 7909090
; TELEFAX: (212) 8609741
; TELE: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 55:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 941 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-179-558-55

Query Match 2.0%; Score 7; DB 4; Length 941;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 308 QRPEEMP 314
DB 102 QRPEEMP 108

RESULT 15
US-08-222-616-33
; Sequence 33, Application US/08222616
; Patent No. 5635177
; GENERAL INFORMATION:
; APPLICANT: Bennett, Brian D.
; APPLICANT: Goeddel, David
; APPLICANT: Lee, James M.
; APPLICANT: Matthews, William
; APPLICANT: Tsai, Siao Ping
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: PROTEIN TYROSINE KINASE AGONIST
; TITLE OF INVENTION: ANTIBODIES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/222,616
; FILING DATE: 4-APR-1994
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/00586
```

; FILING DATE: 22-JAN-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/826935  
; FILING DATE: 22-JAN-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lee, Wendy M.  
; REGISTRATION NUMBER:  
; REFERENCE/DOCKET NUMBER: 821P2.  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415/225-1994  
; TELEFAX: 415/952-9881  
; TELEX: 910/371-7168  
; INFORMATION FOR SEQ ID NO: 33:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1298 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-222-616-33

Query Match 2.0%; Score 7; DB 1; Length 1298;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 202 SLRRRQ 208  
Db 473 SLRRRQ 479

## RESULT 16

US-08-340-011-2  
; Sequence 2, Application US/08340011  
; Patent No. 5776755  
; GENERAL INFORMATION:  
; APPLICANT: Allitalo, et al.  
; TITLE OF INVENTION: FLT4, A NOVEL RECEPTOR TYROSIN KINASE  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower, 233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION NUMBER: US/08/340,011  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA: 07/959,951  
; APPLICATION NUMBER: 07/959,951  
; FILING DATE: 09-OCT-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meyers, Thomas C.  
; REGISTRATION NUMBER: 36,989  
; REFERENCE/DOCKET NUMBER: 32267  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/474-6300  
; TELEFAX: 312/474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1298 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-340-011-2

Query Match 2.0%; Score 7; DB 1; Length 1298;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 202 SLRRRQ 208  
Db 473 SLRRRQ 479

## RESULT 17

US-08-901-710-2  
; Sequence 2, Application US/08901710  
; Patent No. 6107046  
; GENERAL INFORMATION:  
; APPLICANT: Allitalo, Kari  
; APPLICANT: Aprelikova, Olga  
; APPLICANT: Pajusola, Katri  
; APPLICANT: Armstrong, Elina  
; APPLICANT: Korhonen, Jaana  
; APPLICANT: Kaipainen, Arja  
; APPLICANT: Matikainen, Marja-Terttu  
; TITLE OF INVENTION: FLT4, A RECEPTOR TYROSINE KINASE, AND USES  
; TITLE OF INVENTION: THEREOF  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower, 233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION NUMBER: US/08/901,710  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA: 08/340,011  
; APPLICATION NUMBER: 08/340,011  
; FILING DATE: 14-NOV-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/257,754  
; FILING DATE: 09-JUL-1994  
; PRIOR APPLICATION DATA: 07/959,951  
; APPLICATION NUMBER: 07/959,951  
; FILING DATE: 09-OCT-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gass, David A.  
; REGISTRATION NUMBER: 38,153  
; REFERENCE/DOCKET NUMBER: 28113/33824  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/474-6300  
; TELEFAX: 312/474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1298 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-901-710-2

Query Match 2.0%; Score 7; DB 3; Length 1298;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 202 SLRRRQ 208  
Db 473 SLRRRQ 479

## RESULT 18

PCT-US95-04228-33  
; Sequence 33, Application PC/TUS9504228  
; GENERAL INFORMATION:  
; APPLICANT: Genentech, Inc.  
; APPLICANT: Bennett, Brian D.  
; APPLICANT: Goeddel, David  
; APPLICANT: Lee, James M.  
; APPLICANT: Matthews, William  
; APPLICANT: Tsai, Siao Ping  
; APPLICANT: Wood, William I.  
; TITLE OF INVENTION: PROTEIN TYROSINE KINASE AGONIST ANTIBODIES  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 460 Point San Bruno Blvd  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: patin (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/04228  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/222616  
; FILING DATE: 04-APR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Wendy M. Lee  
; REGISTRATION NUMBER: 00,000  
; REFERENCE/DOCKET NUMBER: 821P3PCT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415/225-1994  
; TELEFAX: 415/952-9881  
; TELEX: 910/371-7168  
; INFORMATION FOR SEQ ID NO: 33:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1298 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
PCT-US95-04228-33

Query Match 2.0%; Score 7; DB 5; Length 1298;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 202 SLRRRQQ 208  
| | | | | | | |  
Db 473 SLRRRQQ 479

RESULT 19  
US-08-874-678-33  
; Sequence 33, Application US/08874678  
; Patent No. 5952199  
; GENERAL INFORMATION:  
; APPLICANT: Davis-Smyth, Terri L.  
; APPLICANT: Chen, Helen H.  
; APPLICANT: Presta, Leonard  
; APPLICANT: Ferrara, Napoleone  
; TITLE OF INVENTION: NOVEL INHIBITORS OF VASCULAR ENDOTHELIAL GROWTH FACTOR  
; NUMBER OF SEQUENCES: 48  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
; STREET: Four Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States

ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/874,678  
; FILING DATE: HEREWITH  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/643,839  
; FILING DATE: 07-MAY-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dreger, Walter H.  
; REGISTRATION NUMBER: 24,190  
; REFERENCE/DOCKET NUMBER: A-63291-1/WHd  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 33:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1362 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-08-874-678-33

Query Match 2.0%; Score 7; DB 2; Length 1362;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 202 SLRRRQQ 208  
| | | | | | | |  
Db 472 SLRRRQQ 478

RESULT 20  
US-08-643-839-33  
; Sequence 33, Application US/08643839  
; Patent No. 6100071  
; GENERAL INFORMATION:  
; APPLICANT: Davis-Smyth, Terri L.  
; APPLICANT: Chen, Helen H.  
; APPLICANT: Presta, Leonard  
; APPLICANT: Ferrara, Napoleone  
; TITLE OF INVENTION: NOVEL INHIBITORS OF VASCULAR ENDOTHELIAL  
; TITLE OF INVENTION: GROWTH FACTOR ACTIVITY, THEIR USES AND PROCESSES FOR THEIR  
; NUMBER OF SEQUENCES: 48  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
; STREET: Four Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/643,839  
; FILING DATE: 07-MAY-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dreger, Walter H.  
; REGISTRATION NUMBER: 24,190  
; REFERENCE/DOCKET NUMBER: A-63291/WHd  
; TELECOMMUNICATION INFORMATION:



; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 33:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1362 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-08-643-839-33

Query Match 2.0%; Score 7; DB 3; Length 1362;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 202 SLRRQO 208  
| | | | | | |  
DB 473 SLRRQO 478

RESULT 21  
US-08-340-011-4  
; Sequence 4, Application US/08340011  
; Patent No. 5776755  
; GENERAL INFORMATION:  
; APPLICANT: Alitalo, et al.  
; TITLE OF INVENTION: FLT4, A NOVEL RECEPTOR TYROSIN KINASE  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower, 233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/340,011  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/959,951  
; FILING DATE: 09-OCT-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meyers, Thomas C.  
; REGISTRATION NUMBER: 36,989  
; REFERENCE/DOCKET NUMBER: 32267  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/474-6300  
; TELEFAX: 312/474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1363 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-340-011-4

Query Match 2.0%; Score 7; DB 1; Length 1363;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 202 SLRRQO 208  
| | | | | | |  
DB 473 SLRRQO 479

RESULT 22  
US-08-874-678-32  
; Sequence 32, Application US/08874678  
; Patent No. 5952199  
; GENERAL INFORMATION:  
; APPLICANT: Davis-Smyth, Terri L.  
; APPLICANT: Chen, Helen H.  
; APPLICANT: Presta, Leonard  
; APPLICANT: Ferrara, Napoleone  
; TITLE OF INVENTION: NOVEL INHIBITORS OF VASCULAR ENDOTHELIAL GROWTH FACTOR  
; TITLE OF INVENTION: ACTIVITY, THEIR USES AND PROCESSES FOR THEIR PRODUCTION  
; NUMBER OF SEQUENCES: 48  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
; STREET: Four Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/874,678  
; FILING DATE: HERewith  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/643,839  
; FILING DATE: 07-MAY-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dreger, Walter H.  
; REGISTRATION NUMBER: 24,190  
; REFERENCE/DOCKET NUMBER: A-63291-1/WHd  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 32:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1363 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-08-874-678-32

Query Match 2.0%; Score 7; DB 2; Length 1363;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 202 SLRRQO 208  
| | | | | | |  
DB 473 SLRRQO 479

RESULT 23  
US-08-643-839-32  
; Sequence 32, Application US/08643839  
; Patent No. 6100071  
; GENERAL INFORMATION:  
; APPLICANT: Davis-Smyth, Terri L.  
; APPLICANT: Chen, Helen H.  
; APPLICANT: Presta, Leonard  
; APPLICANT: Ferrara, Napoleone  
; TITLE OF INVENTION: NOVEL INHIBITORS OF VASCULAR ENDOTHELIAL  
; TITLE OF INVENTION: GROWTH FACTOR ACTIVITY, THEIR USES AND PROCESSES FOR THEIR  
; NUMBER OF SEQUENCES: 48  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
; STREET: Four Embarcadero Center, Suite 3400

; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/643,839  
; FILING DATE: 07-MAY-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dreger, Walter H.  
; REGISTRATION NUMBER: 24,190  
; REFERENCE/DOCKET NUMBER: A-63291/WHd  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 32:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1363 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; US-08-643-839-32

Query Match 2.0%; Score 7; DB 3; Length 1363;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 202 SLRRRQ 208  
Db 473 SLRRRQ 479

RESULT 24  
US-08-901-710-4  
; Sequence 4, Application US/08901710  
; Patent No. 6107046  
; GENERAL INFORMATION:  
; APPLICANT: Alitalo, Kari  
; APPLICANT: Aprelikova, Olga  
; APPLICANT: Pajusola, Katri  
; APPLICANT: Armstrong, Elna  
; APPLICANT: Korhonen, Jaana  
; APPLICANT: Kaipainen, Arja  
; APPLICANT: Matikainen, Marja-Terttu  
; TITLE OF INVENTION: FLT4, A RECEPTOR TYROSINE KINASE, AND USES THEREOF  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower, 233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/901,710  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/340,011  
; FILING DATE: 14-NOV-1994

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/257,754  
; FILING DATE: 09-JUL-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/959,951  
; FILING DATE: 09-OCT-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gass, David A.  
; REGISTRATION NUMBER: 38,153  
; REFERENCE/DOCKET NUMBER: 28113/33824  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/474-6300  
; TELEFAX: 312/474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1363 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-901-710-4  
; Query Match 2.0%; Score 7; DB 3; Length 1363;  
; Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
; Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 202 SLRRRQ 208  
Db 473 SLRRRQ 479  
RESULT 25  
US-08-874-678-34  
; Sequence 34, Application US/08874678  
; Patent No. 5952199  
; GENERAL INFORMATION:  
; APPLICANT: Davis-Smyth, Terri L.  
; APPLICANT: Chen, Helen H.  
; APPLICANT: Presta, Leonard  
; APPLICANT: Ferrara, Napoleone  
; TITLE OF INVENTION: NOVEL INHIBITORS OF VASCULAR ENDOTHELIAL GROWTH FACTOR  
; TITLE OF INVENTION: ACTIVITY, THEIR USES AND PROCESSES FOR THEIR PRODUCTION  
; NUMBER OF SEQUENCES: 48  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert  
; STREET: Four Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/874,678  
; FILING DATE: HEREWITH  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/643,839  
; FILING DATE: 07-MAY-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dreger, Walter H.  
; REGISTRATION NUMBER: 24,190  
; REFERENCE/DOCKET NUMBER: A-63291-1/WHd  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 34:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1368 amino acids

; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-08-874-678-34

Query Match 2.0%; Score 7; DB 2; Length 1368;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 202 SLRRRQ 208  
|||||  
Db 478 SLRRRQ 484

## RESULT 26

US-08-643-839-34  
; Sequence 34, Application US/08643839  
; Patent No. 6100071  
; GENERAL INFORMATION:  
; APPLICANT: Davis-Smyth, Terri L.  
; APPLICANT: Chen, Helen H.  
; APPLICANT: Presta, Leonard  
; APPLICANT: Ferrara, Napoleone  
; TITLE OF INVENTION: NOVEL INHIBITORS OF VASCULAR ENDOTHELIAL  
; TITLE OF INVENTION: GROWTH FACTOR ACTIVITY, THEIR USES AND PROCESSES FOR THEIR  
; TITLE OF INVENTION: PRODUCTION  
; NUMBER OF SEQUENCES: 48  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert  
; STREET: Four Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/643,839  
; FILING DATE: 07-MAY-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dreger, Walter H.  
; REGISTRATION NUMBER: 24,190  
; REFERENCE/DOCKET NUMBER: A-63291/WHH  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 34:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1368 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-08-643-839-34

Query Match 2.0%; Score 7; DB 3; Length 1368;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 202 SLRRRQ 208  
|||||  
Db 478 SLRRRQ 484

## RESULT 27

US-08-491-954-21  
; Sequence 21, Application US/08491954

; Patent No. 6096321  
; GENERAL INFORMATION:  
; APPLICANT: Girardeau, Jean-Pierre  
; APPLICANT: Martin, Christine  
; APPLICANT: Mechin, Marie-Claire  
; APPLICANT: Der Vartanian, Maurice  
; APPLICANT: Bousquet, Francois  
; TITLE OF INVENTION: SUB-UNIT OF CS31A PROTEIN CAPSULE  
; TITLE OF INVENTION: MODIFIED BY AT LEAST ONE HETEROLOGOUS PEPTIDE, CS31A  
; TITLE OF INVENTION: PROTEIN CAPSULE INCLUDING SUCH A SUB-UNIT, AND  
; TITLE OF INVENTION: MICROORGANISMS WHOSE OUTER MEMBRANE CARRIES SUCH  
; TITLE OF INVENTION: SUB-UNITS, AND PROCEDURE FOR OBTAINING AND UTILIZING SUCH  
; TITLE OF INVENTION: SUB-UNITS  
; NUMBER OF SEQUENCES: 115  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: WEISER & ASSOCIATES  
; STREET: 230 South Fifteenth Street, Suite 500  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19102  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/491,954  
; FILING DATE: 16-FEB-1996  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/FR93/01281  
; FILING DATE: 21-DEC-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Weiser, Gerard J.  
; REGISTRATION NUMBER: 19,763  
; REFERENCE/DOCKET NUMBER: 989.6264P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-875-8383  
; TELEFAX: 215-875-8394  
; INFORMATION FOR SEQ ID NO: 21:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 9 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-491-954-21

Query Match 1.7%; Score 6; DB 3; Length 9;  
Best Local Similarity 100.0%; Pred. No. 1.7e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 32 GNGVAL 37  
|||||  
Db 1 GNGVAL 6

## RESULT 28

PCT-US92-03432-3  
; Sequence 3, Application PC/TUS9203432  
; GENERAL INFORMATION:  
; APPLICANT: the Trustees of Boston  
; APPLICANT: University  
; TITLE OF INVENTION: METHOD OF ISOLATING LIGANDS  
; TITLE OF INVENTION: RECEPTORS  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: U.S.A.

```
;
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50Z or 55SX
; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
; SOFTWARE: WordPerfect (Version 5.0)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/03432
; FILING DATE: 19920427
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/693/440
; FILING DATE: April 29, 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul T. Clark
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 04766/006W01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11
; TYPE: AMINO ACID
; TOPOLOGY: linear
; PCT-US92-03432-3

Query Match 1.7%: Score 6; DB 5; Length 11;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 56 NLAVAD 61
Db 3 NLAVAD 8

RESULT 29
US-07-798-223A-2
; Sequence 2, Application US/07798223A
; Patent No. 5422108
; GENERAL INFORMATION:
; APPLICANT: Mirkov, T. Erik
; APPLICANT: Fitzmaurice, Leona Claire
; TITLE OF INVENTION: Protection of Plants Against Pathogens
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fitch, Even, Tabin & Flannery
; STREET: 135 South LaSalle Street, Suite 900
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/798, 223A
; FILING DATE: 19911125
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/762,679
; FILING DATE: 19-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L.
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 51984
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619)552-1311
; TELEFAX: (619)552-0095
; INFORMATION FOR SEQ ID NO: 2:
```

```
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-07-798-223A-2

Query Match 1.7%: Score 6; DB 1; Length 18;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 141 ALVILG 146
Db 3 ALVILG 8

RESULT 30
5187078-2
; Patent No. 5187078
; APPLICANT: OHYA, MASAMI;MIZOGUCHI, JUNZO;ONOZAWA, TAKASHI
; TITLE OF INVENTION: PLASMA-TYPE GLUTATHIONE PEROXIDASE GENE
; AND APPLICATION OF THE SAME
; NUMBER OF SEQUENCES: 24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/540,115
; FILING DATE: 19-JUN-1990
; SEQ ID NO:2:
; LENGTH: 24
; 5187078-2

Query Match 1.7%: Score 6; DB 6; Length 24;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 141 ALVILG 146
Db 1 ALVILG 6

RESULT 31
US-08-383-751A-3
; Sequence 3, Application US/08383751A
; Patent No. 5753516
; GENERAL INFORMATION:
; APPLICANT: Heagy, Wyrta E.
; APPLICANT: Finberg, Robert W.
; TITLE OF INVENTION: Identification and Uses of Opioid
; TITLE OF INVENTION: Receptors
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: US
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/383,751A
; FILING DATE: 03-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Wilson, Mark B.
; REGISTRATION NUMBER: 37,259
; REFERENCE/DOCKET NUMBER: DFCI:001/WIM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; TELEX: 79-0924
```

; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 40 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-383-751A-3

Query Match 1.7%; Score 6; DB 1; Length 40;  
Best Local Similarity 100.0%; Pred. No. 73;  
Matches 6; Conservative 0; Mismatches 0; Indels 0;

Gaps 0;

QY 56 NLAVALD 61

Db 19 NLAVALD 24

## RESULT 32

US-08-637-759B-358  
; Sequence 358, Application US/08637759B  
; Patent No. 5876931  
; GENERAL INFORMATION:

APPLICANT: David William Holden  
; TITLE OF INVENTION: Identification of Genes  
; NUMBER OF SEQUENCES: 501  
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Patrea L. Pabst  
STREET: 2800 One Atlantic Center  
STREET: 1201 West Peachtree Street

CITY: Atlanta

STATE: Georgia

COUNTRY: USA

ZIP: 30309-3450

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/637,759B

FILING DATE: 03-MAY-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/GB95/02875

FILING DATE: 11-DEC-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Pabst, Patrea L.

REGISTRATION NUMBER: 31,284

REFERENCE/DOCKET NUMBER: RPMS 101

TELECOMMUNICATION INFORMATION:

TELEPHONE: (404) 873-8794

TELEFAX: (404) 873-8795

INFORMATION FOR SEQ ID NO: 358:

SEQUENCE CHARACTERISTICS:

LENGTH: 56 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

US-08-637-759B-358

Query Match 1.7%; Score 6; DB 2; Length 56;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0;

Gaps 0;

QY 103 IVFLTV 108

Db 8 IVFLTV 13

## RESULT 33

US-08-871-355A-358  
; Sequence 358, Application US/08871355A  
; Patent No. 6015669  
; GENERAL INFORMATION:

APPLICANT: David William Holden

TITLE OF INVENTION: Identification of Genes

NUMBER OF SEQUENCES: 501

CORRESPONDENCE ADDRESS:

ADDRESSEE: Patrea L. Pabst

STREET: 2800 One Atlantic Center

STREET: 1201 West Peachtree Street

CITY: Atlanta

STATE: Georgia

COUNTRY: USA

ZIP: 30309-3450

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/871,355A

FILING DATE: 09-JUN-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/GB95/02875

FILING DATE: 11-DEC-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Pabst, Patrea L.

REGISTRATION NUMBER: 31,284

REFERENCE/DOCKET NUMBER: RPMS 101 CON

TELECOMMUNICATION INFORMATION:

TELEPHONE: (404) 873-8794

TELEFAX: (404) 873-8795

INFORMATION FOR SEQ ID NO: 358:

SEQUENCE CHARACTERISTICS:

LENGTH: 56 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

US-08-871-355A-358

## Query Match

1.7%; Score 6; DB 3; Length 56;

Best Local Similarity 100.0%; Pred. No. 1e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0;

Gaps 0;

QY 103 IVFLTV 108

Db 8 IVFLTV 13

## RESULT 34

US-09-201-945-358

; Sequence 358, Application US/09201945

; Patent No. 6342215

; GENERAL INFORMATION:

APPLICANT: David William Holden

TITLE OF INVENTION: Identification of Genes

NUMBER OF SEQUENCES: 501

CORRESPONDENCE ADDRESS:

ADDRESSEE: Patrea L. Pabst

STREET: 2800 One Atlantic Center

STREET: 1201 West Peachtree Street

CITY: Atlanta

STATE: Georgia

COUNTRY: USA

ZIP: 30309-3450

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

```
;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/201,945
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/637,759
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: RPS 101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 873-8794
; TELEFAX: (404) 873-8795
; INFORMATION FOR SEQ ID NO: 358:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 56 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; US-09-201-945-358

Query Match 1.7%; Score 6; DB 4; Length 56;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 103 IVFLTV 108
Db 8 IVFLTV 13

RESULT 35
US-08-513-974B-34
; Sequence 34, Application US/08513974B
; Patent No. 6114139
; GENERAL INFORMATION:
; APPLICANT: Hinuma, Shuji
; APPLICANT: Hosoya, Masaki
; APPLICANT: Fujii, Ryo
; APPLICANT: Ohtaki, Tetsuya
; APPLICANT: Fukusumi, Shoji
; APPLICANT: Ohgi, Kazuhiro
; TITLE OF INVENTION: G PROTEIN COUPLED RECEPTOR PROTEIN,
; TITLE OF INVENTION: PRODUCTION, AND USE THEREOF
; NUMBER OF SEQUENCES: 380
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/513,974B
; FILING DATE: 14-SEP-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP95/01599
; FILING DATE: 10-AUG-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 7-093989
; FILING DATE: 19-AUG-1995
```

```
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 7-057186
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 7-007177
; FILING DATE: 20-JAN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 6-326611
; FILING DATE: 28-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 6-270017
; FILING DATE: 02-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 6-236357
; FILING DATE: 30-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 6-236356
; FILING DATE: 30-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 6-189274
; FILING DATE: 11-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 6-189273
; FILING DATE: 11-AUG-1945
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 6-189272
; FILING DATE: 11-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Resnick, David S.
; REGISTRATION NUMBER: 34,235
; REFERENCE/DOCKET NUMBER: 45753
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 70 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-513-974B-34

Query Match 1.7%; Score 6; DB 3; Length 70;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 56 NLAADV 61
Db 21 NLAADV 26

RESULT 36
US-08-513-974B-317
; Sequence 317, Application US/08513974B
; Patent No. 6114139
; GENERAL INFORMATION:
; APPLICANT: Hinuma, Shuji
; APPLICANT: Hosoya, Masaki
; APPLICANT: Fujii, Ryo
; APPLICANT: Ohtaki, Tetsuya
; APPLICANT: Fukusumi, Shoji
; APPLICANT: Ohgi, Kazuhiro
; TITLE OF INVENTION: G PROTEIN COUPLED RECEPTOR PROTEIN,
; TITLE OF INVENTION: PRODUCTION, AND USE THEREOF
; NUMBER OF SEQUENCES: 380
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
```

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/513,974B  
FILING DATE: 14-SEP-1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP95/01599  
FILING DATE: 10-AUG-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 7-093989  
FILING DATE: 19-AUG-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 7-057186  
FILING DATE: 16-MAR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 7-007177  
FILING DATE: 20-JAN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-326611  
FILING DATE: 28-DEC-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-270017  
FILING DATE: 02-NOV-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-236357  
FILING DATE: 30-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-236356  
FILING DATE: 30-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-189274  
FILING DATE: 11-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-189273  
FILING DATE: 11-AUG-1945  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-189272  
FILING DATE: 11-AUG-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Resnick, David S.  
REGISTRATION NUMBER: 34,235  
REFERENCE/DOCKET NUMBER: 45753  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-523-3400  
TELEFAX: 617-523-6440  
INFORMATION FOR SEQ ID NO: 317:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 70 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-513-974B-317

Query Match 1.7%; Score 6; DB 3; Length 70;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 NLAVALD 61  
Db 21 NLAVALD 26

RESULT 37  
US-07-971-160-14  
Sequence 14, Application US/07971160  
Patent No. 5474896  
GENERAL INFORMATION:  
APPLICANT: DuJon, Bernard  
APPLICANT: Choulika, Andre

APPLICANT: Colleaux, Laurence  
APPLICANT: Fairhead, Cecile  
APPLICANT: Perrin, Arnaud  
APPLICANT: Plessis, Anne  
APPLICANT: Thierry, Agnes  
TITLE OF INVENTION: Nucleotide Sequence Encoding the Enzyme  
TITLE OF INVENTION: I-SCEI and the Uses Thereof  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
ADDRESSEE: Dunner  
STREET: 1300 I Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/971,160  
FILING DATE: 19921105  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/879,689  
FILING DATE: 05-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Ogden, Stasia L.  
REGISTRATION NUMBER: 36,228  
REFERENCE/DOCKET NUMBER: 03495-0111-01000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 80 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-07-971-160-14

Query Match 1.7%; Score 6; DB 1; Length 80;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 143 VILGTV 148  
Db 9 VILGTV 14

RESULT 38  
US-08-336-241-14  
Sequence 14, Application US/08336241  
Patent No. 5792632  
GENERAL INFORMATION:  
APPLICANT: Choulika, Andre  
APPLICANT: Perrin, Arnaud  
APPLICANT: DuJon, Bernard  
APPLICANT: Nicolas, Jean-Francois  
TITLE OF INVENTION: Nucleotide Sequence Encoding the Enzyme  
TITLE OF INVENTION: I-SCEI and the Uses Thereof  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
ADDRESSEE: Dunner  
STREET: 1300 I Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315

```
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/336,241
; FILING DATE: 07-NOV-1994
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/971,160
; FILING DATE: 05-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/879,689
; FILING DATE: 05-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Potter, Jane E.R.
; REGISTRATION NUMBER: 33,332
; REFERENCE/DOCKET NUMBER: 03495-0111-03000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4400
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 80 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-336-241-14

Query Match 1.7% Score 6; DB 1; Length 80;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 143 VILGTV 148
Db 9 VILGTV 14

RESULT 39
; Sequence 14, Application US/08465273
; Patent No. 5866361
; GENERAL INFORMATION:
; APPLICANT: Choulika, Andre
; APPLICANT: Perrin, Arnaud
; APPLICANT: Dujon, Bernard
; APPLICANT: Nicolas, Jean-Francois
; TITLE OF INVENTION: Nucleotide Sequence Encoding the Enzyme
; Patent No. 5866361
; TITLE OF INVENTION: I-SCEI and the Uses Thereof
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,273
; FILING DATE: 06-MAY-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/336,241
; FILING DATE: 07-NOV-1994
; CLASSIFICATION: 800
```

```
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/971,160
; FILING DATE: 05-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/879,689
; FILING DATE: 05-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Potter, Jane E.R.
; REGISTRATION NUMBER: 33,332
; REFERENCE/DOCKET NUMBER: 03495-0111-06000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 80 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-465-273-14

Query Match 1.7% Score 6; DB 2; Length 80;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 143 VILGTV 148
Db 9 VILGTV 14

RESULT 40
; US-09-119-024-14
; Sequence 14, Application US/09119024
; Patent No. 5948678
; GENERAL INFORMATION:
; APPLICANT: Choulika, Andre
; APPLICANT: Perrin, Arnaud
; APPLICANT: Dujon, Bernard
; APPLICANT: Nicolas, Jean-Francois
; TITLE OF INVENTION: Nucleotide Sequence Encoding the Enzyme
; Patent No. 5948678
; TITLE OF INVENTION: I-SCEI and the Uses Thereof
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/119,024
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/336,241
; FILING DATE: 07-NOV-1994
; APPLICATION NUMBER: US 07/971,160
; FILING DATE: 05-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/879,689
; FILING DATE: 05-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Potter, Jane E.R.
; REGISTRATION NUMBER: 33,332
; REFERENCE/DOCKET NUMBER: 03495-0111-03000
; TELECOMMUNICATION INFORMATION:
```



TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 80 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-119-024-14

Query Match 1.7%; Score 6; DB 2; Length 80;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0;

Qy 143 VILGTV 148  
Db 9 VILGTV 14

RESULT 41  
US-08-417-226-14  
; Sequence 14, Application US/08417226  
; Patent No. 5962327  
; GENERAL INFORMATION:  
; APPLICANT: Choulika, Andre  
; APPLICANT: Perrin, Arnaud  
; APPLICANT: Dujon, Bernard  
; APPLICANT: NICOLAS, Jean-Francois  
; TITLE OF INVENTION: Nucleotide Sequence Encoding the Enzyme  
; Patent No. 5962327  
; TITLE OF INVENTION: I-SCEI and the Uses Thereof  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
; ADDRESSEE: Dunner  
; STREET: 1300 I Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3315  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/417,226  
; FILING DATE: 05-APRIL-1995  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/336,241  
; FILING DATE: 07-NOV-1994  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/971,160  
; FILING DATE: 05-NOV-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/879,689  
; FILING DATE: 05-MAY-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Potter, Jane E.R.  
; REGISTRATION NUMBER: 33,332  
; REFERENCE/DOCKET NUMBER: 03495-0111-04000  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-408-4000  
; TELEFAX: 202-408-4400  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 80 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-417-226-14

Query Match 1.7%; Score 6; DB 2; Length 80;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0;

Qy 143 VILGTV 148  
Db 9 VILGTV 14

RESULT 42  
US-09-196-131-14  
; Sequence 14, Application US/09196131  
; Patent No. 6238924  
; GENERAL INFORMATION:  
; APPLICANT: Choulika, Andre  
; APPLICANT: Perrin, Arnaud  
; APPLICANT: Dujon, Bernard  
; APPLICANT: NICOLAS, Jean-Francois  
; TITLE OF INVENTION: Nucleotide Sequence Encoding the Enzyme  
; Patent No. 6238924  
; TITLE OF INVENTION: I-SCEI and the Uses Thereof  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
; ADDRESSEE: Dunner  
; STREET: 1300 I Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3315  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/196,131  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/417,226  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/879,689  
; FILING DATE: 05-MAY-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Potter, Jane E.R.  
; REGISTRATION NUMBER: 33,332  
; REFERENCE/DOCKET NUMBER: 03495-0111-04000  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-408-4000  
; TELEFAX: 202-408-4400  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 80 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-09-196-131-14

Query Match 1.7%; Score 6; DB 4; Length 80;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0;

Qy 143 VILGTV 148  
Db 9 VILGTV 14

RESULT 43  
US-08-961-083-130

```
; Sequence 130, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; PRIOR APPLICATION DATA:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 130:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 82 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-961-083-130

Query Match 1.7%; Score 6; DB 4; Length 82;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 113 RYFKVV 118
Db 46 RYFKVV 51

RESULT 44
US-08-091-569-21
; Sequence 21, Application US/08091569
; Patent No. 5494792
; GENERAL INFORMATION:
; APPLICANT: KAMBOJ, Rajender
; APPLICANT: NUTT, Stephen
; APPLICANT: SHEKTER, Lee
; APPLICANT: WOSNICK, Michael
; TITLE OF INVENTION: KAINATE-BINDING HUMAN CNS RECEPTORS OF
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/203,676
; FILING DATE:
; PRIOR APPLICATION DATA:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 16777/149/ALLE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)836-9300
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 21:
```

```
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/091,569
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/750,081
; FILING DATE: 27-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 16777/149/ALLE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)836-9300
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 88 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-091-569-21

Query Match 1.7%; Score 6; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 LLIVAF 26
Db 8 LLIVAF 13

RESULT 45
US-08-203-676-21
; Sequence 21, Application US/08203676
; Patent No. 5614406
; GENERAL INFORMATION:
; APPLICANT: KAMBOJ, Rajender
; APPLICANT: NUTT, Stephen
; APPLICANT: SHEKTER, Lee
; APPLICANT: WOSNICK, Michael
; TITLE OF INVENTION: KAINATE-BINDING HUMAN CNS RECEPTORS OF
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/203,676
; FILING DATE:
; PRIOR APPLICATION DATA:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 16777/149/ALLE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)836-9300
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 21:
```

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 88 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-08-203-676-21

Query Match 1.7%; Score 6; DB 1; Length 88;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 21 LLIVAF 26  
| | | | |  
Db 8 LLIVAF 13

Search completed: October 30, 2002, 18:27:56  
Job time : 19 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2002, 18:26:03 ; Search time 33 seconds

(without alignments)  
1164.592 Million cell updates/sec

Title: US-09-886-041-2

Perfect score: 346

Sequence: 1 MTNGSCRIEGTISOVMP.....ANSFQSDQWDPHIVNH 346

Scoring table:

OLIGO

Gapop 60.0 , Gapert 60.0

Searched: 747574 seqs, 111073796 residues

Word size : 0

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

A\_Geneseq\_032802.\*

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	346	100.0	346	22	ABB44522 Human GPCR1a polyp
2	346	100.0	346	22	AAU06197 Novel human G prot
3	346	100.0	346	22	AAU04373 Human G-protein co
4	346	100.0	346	22	AAU080968 Human ngPCR11 #2.
5	296	85.5	296	22	AAU080934 Human ngPCR11 #1.
6	234	67.6	346	22	ABB44523 Human GPCR1c polyp
7	14	4.0	363	20	AAU94654 G-protein coupled
8	14	4.0	363	20	AAU04379 Human G-protein co
9	14	4.0	387	21	AAU90637 Human G protein-co
10	14	4.0	387	21	AAU90672 Human mutant G pro
11	9	2.6	372	21	AAU08621 A human P2Y-like 7

12	9	2.6	372	21	AAU71292 Human orphan G pro
13	9	2.6	372	21	AAU02826 Human G protein co
14	9	2.6	372	21	AAU79564 Human G protein co
15	9	2.6	372	22	AAU64291 Human GTP-binding
16	9	2.6	372	22	AAU02500 Human CON217 G pro
17	9	2.6	372	22	AAU81125 Human G-protein co
18	9	2.6	372	22	AAU72590 Human G-protein co
19	9	2.6	372	22	AAU46598 Human p2Y11 protel
20	9	2.6	372	22	AAU61612 Human protein HP03
21	9	2.6	391	22	AAU03817 G protein-coupled
22	9	2.6	422	22	AAU04547 Human G-protein co
23	8	2.3	320	21	AAU77734 Seven times membra
24	8	2.3	350	22	AAU56337 Non-endogenous hum
25	8	2.3	372	20	AAU86323 Kidney injury asso
26	7	2.0	9	18	AAU13835 Flt4/IgG (80 kDa)
27	7	2.0	11	22	AAU68711 Human Chk2 kinase
28	7	2.0	11	22	AAU69159 Human Chk2 kinase
29	7	2.0	11	22	AAU69607 Human Chk2 kinase
30	7	2.0	11	22	AAU70055 Human Chk2 kinase
31	7	2.0	26	22	AAU58297 Human brain expres
32	7	2.0	32	21	AAU33861 Human secreted pro
33	7	2.0	52	22	AAU70813 Human bone marrow
34	7	2.0	53	22	AAU80438 Human haematologic
35	7	2.0	59	22	AAU17871 Human nervous syst
36	7	2.0	59	22	AAU80364 Human haematologic
37	7	2.0	61	22	AAU80455 Human haematologic
38	7	2.0	62	22	AAU80530 Human haematologic
39	7	2.0	121	22	AAU80402 Human haematologic
40	7	2.0	125	22	AAU19711 Human novel extrac
41	7	2.0	174	19	AAU11057 H. pylori ORF hp2e
42	7	2.0	216	22	AAU69822 Drosophila melanog
43	7	2.0	242	21	AAU42660 Human ORFX ORF2424
44	7	2.0	268	22	AAU19719 Human novel extrac
45	7	2.0	271	22	AAU15980 Human novel secret

ALIGNMENTS

RESULT 1  
ABB44522  
ID ABB44522 standard; Protein; 346 AA.  
XX  
AC ABB44522;  
XX  
DT 28-JAN-2002 (first entry)  
XX  
DE Human GPCR1a polypeptide SEQ ID NO 2.  
XX  
KW Human; GPCR; G-coupled protein-receptor; cardiant; antiarteriosclerotic;  
KW anabolic; cytosstatic; antiviral; gene therapy; cardiomyopathy; obesity;  
KW anorexia; diabetes; osteoporosis; Crohn's disease; multiple sclerosis;  
KW asthma; Alzheimer's disease; Parkinson's disorder; Huntington's disease;  
KW Infection; human immunodeficiency virus; HIV.  
XX  
OS Homo sapiens.  
XX  
PN WO200174904-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US10241.  
XX  
PR 31-MAR-2000; 2000US-193664P.  
PR 03-APR-2000; 2000US-194614P.  
PR 06-APR-2000; 2000US-195063P.  
PR 06-APR-2000; 2000US-195066P.  
PR 06-APR-2000; 2000US-195067P.  
PR 06-APR-2000; 2000US-195068P.  
PR 06-APR-2000; 2000US-195069P.  
PR 06-APR-2000; 2000US-195070P.  
PR 06-APR-2000; 2000US-195510P.  
PR 21-JUL-2000; 2000US-219855P.



```
Query Match      100.0%; Score 346; DB 22; Length 346;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYNGSCCRIEGDTISQVMPPELLIVAFVLGALGNGVALCGCFHMKTKWKPSTVYLFNLAVA 60
Db 1 MYNGSCCRIEGDTISQVMPPELLIVAFVLGALGNGVALCGCFHMKTKWKPSTVYLFNLAVA 60
QY 61 DFLMICLPERTDYLLRRRHWAEGDIPCRVGLFTLAMNRAGSIVFLTVVAADRYFKVHP 120
Db 61 DFLMICLPERTDYLLRRRHWAEGDIPCRVGLFTLAMNRAGSIVFLTVVAADRYFKVHP 120
QY 121 HHAVNTISTRVAAGIVCTLWALVILGTVYLLLENHLCVQETAVSCSFIMESANGWHDIM 180
Db 121 HHAVNTISTRVAAGIVCTLWALVILGTVYLLLENHLCVQETAVSCSFIMESANGWHDIM 180
QY 181 FQLEFFMPLGIILFCSPKIWSLRRRQOLARQARMKKATRFIMVVAIVFITCYLPSVSAR 240
Db 181 FQLEFFMPLGIILFCSPKIWSLRRRQOLARQARMKKATRFIMVVAIVFITCYLPSVSAR 240
QY 241 LYFLWTVPPSACDPSVHGALHITLSFTYMNMLDPLVYFSSPFPKFKYKTKICSLKPK 300
Db 241 LYFLWTVPPSACDPSVHGALHITLSFTYMNMLDPLVYFSSPFPKFKYKTKICSLKPK 300
QY 301 QPGHKTQRPPEEMPISNLGRRCISVANSFQSDGQWDPHIVEWH 346
Db 301 QPGHKTQRPPEEMPISNLGRRCISVANSFQSDGQWDPHIVEWH 346

RESULT 3
AAU04373
ID AAU04373 standard; Protein; 346 AA.
XX
AC AAU04373;
XX
Dt 23-OCT-2001 (first entry)
XX
DE Human G-protein coupled receptor, hRUP19.
XX
KW Human; G-protein coupled receptor; GPCR; hRUP19; agonist;
KW Inverse agonist; lung cancer.
XX
OS Homo sapiens.
XX
PN WO200136471-A2.
XX
PD 25-MAY-2001.
XX
PF 16-NOV-2000; 2000WO-US31509.
XX
PR 17-NOV-1999; 99US-0166088.
PR 17-NOV-1999; 99US-0166099.
PR 17-NOV-1999; 99US-0166369.
PR 23-DEC-1999; 99US-0171900.
PR 23-DEC-1999; 99US-0171901.
PR 23-DEC-1999; 99US-0171902.
PR 11-FEB-2000; 2000US-0181749.
PR 14-MAR-2000; 2000US-0189258.
PR 14-MAR-2000; 2000US-0189259.
PR 10-APR-2000; 2000US-0195898.
PR 10-APR-2000; 2000US-0195899.
PR 10-APR-2000; 2000US-0196078.
PR 28-APR-2000; 2000US-0200419.
PR 12-MAY-2000; 2000US-0203630.
PR 12-JUN-2000; 2000US-0210741.
PR 12-JUN-2000; 2000US-0210982.
PR 21-AUG-2000; 2000US-0226760.
PR 26-SEP-2000; 2000US-0235418.
PR 26-SEP-2000; 2000US-0235779.
PR 20-OCT-2000; 2000US-0242332.
PR 20-OCT-2000; 2000US-0242343.
XX
```

```
(AREN-) ARENA PHARM INC.
XX
PI Chen R, Dang HT, Lowitz KP;
XX
DR WPI; 2001-355616/37.
XX N-PSDB; AAS07946.
XX
PT Endogenous and non-endogenous versions of human G-protein coupled
PT receptors for direct identification of candidate compounds as agonists,
PT inverse agonists or partial agonists for use as therapeutic agents -
XX
PS Claim 45; Page 110-111; 160pp; English.
XX
CC The sequence represents a human G-protein coupled receptor (GPCR),
CC hRUP19. The endogenous and non-endogenous, constitutively activated
CC versions of human G-protein coupled receptors (GPCR), are useful for
CC direct identification of candidate compounds as receptor agonists,
CC inverse agonists or partial agonists having applicability as therapeutic
CC agents for treating diseases related to GPCR, e.g. lung cancer.
CC Non-endogenous version of human GPCRs are also utilized in research
CC settings and in vitro and in vivo system, incorporating GPCRs can be
CC utilised to elucidate and understand the roles these receptors
CC play in the human condition, both normal and diseased.
XX
SQ Sequence 346 AA;
XX
Query Match      100.0%; Score 346; DB 22; Length 346;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYNGSCCRIEGDTISQVMPPELLIVAFVLGALGNGVALCGCFHMKTKWKPSTVYLFNLAVA 60
Db 1 MYNGSCCRIEGDTISQVMPPELLIVAFVLGALGNGVALCGCFHMKTKWKPSTVYLFNLAVA 60
QY 61 DFLMICLPERTDYLLRRRHWAEGDIPCRVGLFTLAMNRAGSIVFLTVVAADRYFKVHP 120
Db 61 DFLMICLPERTDYLLRRRHWAEGDIPCRVGLFTLAMNRAGSIVFLTVVAADRYFKVHP 120
QY 121 HHAVNTISTRVAAGIVCTLWALVILGTVYLLLENHLCVQETAVSCSFIMESANGWHDIM 180
Db 121 HHAVNTISTRVAAGIVCTLWALVILGTVYLLLENHLCVQETAVSCSFIMESANGWHDIM 180
QY 181 FQLEFFMPLGIILFCSPKIWSLRRRQOLARQARMKKATRFIMVVAIVFITCYLPSVSAR 240
Db 181 FQLEFFMPLGIILFCSPKIWSLRRRQOLARQARMKKATRFIMVVAIVFITCYLPSVSAR 240
QY 241 LYFLWTVPPSACDPSVHGALHITLSFTYMNMLDPLVYFSSPFPKFKYKTKICSLKPK 300
Db 241 LYFLWTVPPSACDPSVHGALHITLSFTYMNMLDPLVYFSSPFPKFKYKTKICSLKPK 300
QY 301 QPGHKTQRPPEEMPISNLGRRCISVANSFQSDGQWDPHIVEWH 346
Db 301 QPGHKTQRPPEEMPISNLGRRCISVANSFQSDGQWDPHIVEWH 346

RESULT 4
AAG80968
ID AAG80968 standard; Protein; 346 AA.
XX
AC AAG80968;
XX
DT 28-AUG-2001 (first entry)
XX
DE Human nGPCR11 #2.
XX
KW G protein-coupled receptor; nGPCR; seven transmembrane receptor;
KW signal transduction; schizophrenia; thyroid disorder; renal failure;
KW rheumatoid arthritis; CNS disorder; infection; metabolic disease;
KW cardiovascular disease; proliferative disorder; hormonal disorder;
KW neurological disorder; neuronal disorder; Alzheimer's disease; cancer;
KW attention deficit-hyperactivity disorder/attention deficit disorder;
KW Parkinson's disease; migraine; senile dementia; inflammatory disease;
KW rheumatoid arthritis; autoimmune disorder; respiratory ailment;
```

KW neuroprotective.  
 OS Homo sapiens.  
 XX WO200136473-A2.  
 PN  
 XX  
 PD 25-MAY-2001.  
 XX  
 PF 16-NOV-2000; 2000WO-US31581.  
 XX  
 PR 16-NOV-1999; 99US-0165838.  
 PR 17-NOV-1999; 99US-0166071.  
 PR 19-NOV-1999; 99US-0166678.  
 PR 28-DEC-1999; 99US-0173396.  
 PR 22-FEB-2000; 2000US-0184129.  
 PR 28-FEB-2000; 2000US-0185421.  
 PR 28-FEB-2000; 2000US-0185554.  
 PR 02-MAR-2000; 2000US-0186530.  
 PR 03-MAR-2000; 2000US-0186811.  
 PR 09-MAR-2000; 2000US-0188114.  
 PR 17-MAR-2000; 2000US-0190310.  
 PR 21-MAR-2000; 2000US-0190800.  
 PR 20-APR-2000; 2000US-0198568.  
 PR 02-MAY-2000; 2000US-0201190.  
 PR 08-MAY-2000; 2000US-0203111.  
 PR 25-MAY-2000; 2000US-0207094.  
 XX  
 PA (PHAA ) PHARMACIA & UPJOHN CO.  
 XX  
 PI Vogeli G, Wood LS, Parodi LA, Hiebsch RR, Lind P, Slightom J;  
 PI Schellin KA, Kaytes PS, Bannigan CM, Ruff V, Sejltz T, Huff RM;  
 XX WPI: 2001-389826/41.  
 DR N-PSDB: AAH51008.  
 XX  
 XX New G protein-coupled receptor (ngPCR-x) and its encoding  
 PT polynucleotide useful for diagnosing and treating e.g. schizophrenia -  
 PS Claim 37: Page 89; 261pp; English.  
 XX  
 CC The present invention relates to novel G protein-coupled receptors  
 CC (ngPCRx; where x is 1, 3, 4, 5, 9, 11, 12, 14-18, 20, 21, 22, 24, 27,  
 CC 28, 31-38, 40, 41, 53-60) and their coding sequences. The present  
 CC sequence is one such G protein-coupled receptor. GPCR's are also known as  
 CC seven transmembrane receptors and function in signal transduction. The  
 CC ngPCRx coding sequences are useful for screening a human to diagnose a  
 CC disorder affecting the brain or a genetic predisposition, specifically  
 CC schizophrenia. ngPCRx are useful for identifying compounds useful for  
 CC treating schizophrenia. Detection of ngPCRx in a sample is useful as a  
 CC diagnostic tool for diseases or disorders e.g. thyroid disorders, renal  
 CC failure, rheumatoid arthritis, CNS disorders, infections such as HIV-1,  
 CC metabolic and cardiovascular diseases, proliferative disorders and  
 CC hormonal disorders. Modulators of ngPCRx activity have the utility for  
 CC treating neurological disorders, including schizophrenia, ADHD/ADD  
 CC (attention deficit-hyperactivity disorder/attention deficit disorder),  
 CC and neuronal disorders such as Alzheimer's disease, Parkinson's disease,  
 CC migraine and senile dementia. Additional disorders include inflammatory  
 CC conditions (e.g. Crohn's disease), rheumatoid arthritis, autoimmune  
 CC disorders, cancers, respiratory ailments such as asthma, and inflammatory  
 CC diseases e.g. inflammatory bowel disease.  
 XX  
 SQ Sequence 346 AA;  
 Query Match 100.0%; Score 346; DB 22; Length 346;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 MYNSCCRIEGDTISQVMPPLIIAFAVLGALNGVALCGFCFHKMTKPSIVYLFNLAVA 60  
 Db 1 MYNSCCRIEGDTISQVMPPLIIAFAVLGALNGVALCGFCFHKMTKPSIVYLFNLAVA 60  
 Qy 61 DFLLMICLPFRDYLLRRRHWAFCGDI PCRVGLFTLMMNRAGSI VFLTVVAADRFKVVHP 120  
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 61 DFLLMICLPFRDYLLRRRHWAFCGDI PCRVGLFTLMMNRAGSI VFLTVVAADRFKVVHP 120  
 Qy 121 HHAVNTISTRVAAGIVCTLMALVILGTIVYLLLENHLCVQETAVSCSFIMESANGWHDIM 180  
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 121 HHAVNTISTRVAAGIVCTLMALVILGTIVYLLLENHLCVQETAVSCSFIMESANGWHDIM 180  
 Qy 181 FOLEFFMPPLGILFCSPKIVWSLRRRQLARQARMKATREIMVVAIVFITCYLPSVSAR 240  
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 181 FOLEFFMPPLGILFCSPKIVWSLRRRQLARQARMKATREIMVVAIVFITCYLPSVSAR 240  
 Qy 241 LYFLWTPSSACDPVSGALHITLSFTYMSMLDPLVYFSSPSPFPKFNKIKICSLKPK 300  
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 241 LYFLWTPSSACDPVSGALHITLSFTYMSMLDPLVYFSSPSPFPKFNKIKICSLKPK 300  
 Qy 301 QPGHKTQRPPEMPISNLGRSCISVANSFQSDGQWDPHIVEWH 346  
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 301 QPGHKTQRPPEMPISNLGRSCISVANSFQSDGQWDPHIVEWH 346  
 RESULT 5  
 AAG80934  
 ID AAG80934 standard; Protein; 296 AA.  
 XX  
 AC AAG80934;  
 XX  
 DT 28-AUG-2001 (first entry)  
 XX  
 DE Human ngPCR11 #1.  
 XX  
 KW G protein-coupled receptor; ngPCR; seven transmembrane receptor;  
 KW signal transduction; schizophrenia; thyroid disorder; renal failure;  
 KW rheumatoid arthritis; CNS disorder; infection; metabolic disease;  
 KW cardiovascular disease; proliferative disorder; hormonal disorder;  
 KW neurological disorder; neuronal disorder; Alzheimer's disease; cancer;  
 KW attention deficit-hyperactivity disorder/attention deficit disorder;  
 KW Parkinson's disease; migraine; senile dementia; inflammatory disease;  
 KW rheumatoid arthritis; autoimmune disorder; respiratory ailment;  
 KW neuroprotective.  
 OS Homo sapiens.  
 XX  
 PN WO200136473-A2.  
 XX  
 PD 25-MAY-2001.  
 XX  
 PF 16-NOV-2000; 2000WO-US31581.  
 XX  
 PR 16-NOV-1999; 99US-0165838.  
 PR 17-NOV-1999; 99US-0166071.  
 PR 19-NOV-1999; 99US-0166678.  
 PR 28-DEC-1999; 99US-0173396.  
 PR 22-FEB-2000; 2000US-0184129.  
 PR 28-FEB-2000; 2000US-0185421.  
 PR 28-FEB-2000; 2000US-0185554.  
 PR 02-MAR-2000; 2000US-0186530.  
 PR 03-MAR-2000; 2000US-0186811.  
 PR 09-MAR-2000; 2000US-0190310.  
 PR 17-MAR-2000; 2000US-0190800.  
 PR 21-MAR-2000; 2000US-0190800.  
 PR 20-APR-2000; 2000US-0198568.  
 PR 02-MAY-2000; 2000US-0201190.  
 PR 08-MAY-2000; 2000US-0203111.  
 PR 25-MAY-2000; 2000US-0207094.  
 XX  
 PA (PHAA ) PHARMACIA & UPJOHN CO.  
 XX  
 PI Vogeli G, Wood LS, Parodi LA, Hiebsch RR, Lind P, Slightom J;  
 PI Schellin KA, Kaytes PS, Bannigan CM, Ruff V, Sejltz T, Huff RM;  
 XX WPI: 2001-389826/41.  
 DR N-PSDB: AAH50974.  
 XX  
 PT New G protein-coupled receptor (ngPCR-x) and its encoding



PT polynucleotide useful for diagnosing and treating e.g. schizophrenia -  
PS Claim 37; Pages 77-78; 261pp; English.  
XX The present invention relates to novel G protein-coupled receptors  
XX (GPCRx; where x is 1, 3, 4, 5, 9, 11, 12, 14-18, 20, 21, 22, 24, 27,  
CC 28, 31-38, 40, 41, 53-60) and their coding sequences. The present  
CC sequence is one such G protein-coupled receptor. GPCRs are also known as  
CC seven transmembrane receptors and function in signal transduction. The  
CC nGPCRx coding sequences are useful for screening a human to diagnose a  
CC disorder affecting the brain or a genetic predisposition, specifically  
CC schizophrenia. nGPCRx are useful for identifying compounds useful for  
CC treating schizophrenia. Detection of nGPCRx in a sample is useful as a  
CC diagnostic tool for diseases or disorders e.g. thyroid disorders, renal  
CC failure, rheumatoid arthritis, CNS disorders, infections such as HIV-1,  
CC metabolic and cardiovascular diseases, proliferative disorders and  
CC hormonal disorders. Modulators of nGPCRx activity have the utility for  
CC treating neurological disorders, including schizophrenia, ADHD/ADD  
CC (attention deficit-hyperactivity disorder/attention deficit disorder),  
CC and neuronal disorders such as Alzheimer's disease, Parkinson's disease,  
CC migraine and senile dementia. Additional disorders include inflammatory  
CC conditions (e.g. Crohn's disease), rheumatoid arthritis, autoimmune  
CC disorders, cancers, respiratory ailments such as asthma, and inflammatory  
CC diseases e.g. inflammatory bowel disease.

SQ Sequence 296 AA;  
Query Match 85.5%; Score 296; DB 22; Length 296;  
Best Local Similarity 100.0%; Pred. No. 6.9e-303; Indels 0; Gaps 0;  
Matches 296; Conservative 0; Mismatches 0;

QY 21 LLVAFVLGALGVGVALGCGFHKMTKPTVYLFNLAVADFLMCLPRTDYLLRRH 80  
Db 1 LLVAFVLGALGVGVALGCGFHKMTKPTVYLFNLAVADFLMCLPRTDYLLRRH 60  
QY 81 WAFGDIPIRVLGTLTANRAGSIVLTVVAADRYFKVYVHPHVAVTISTRVAAGIVCTLW 140  
Db 61 WAFGDIPIRVLGTLTANRAGSIVLTVVAADRYFKVYVHPHVAVTISTRVAAGIVCTLW 120  
QY 141 ALVILGVYVLLLENHLCVQETAVSCSEFIMESANGWHDIMFQLEFFMPLGILIFCSFKIV 200  
Db 121 ALVILGVYVLLLENHLCVQETAVSCSEFIMESANGWHDIMFQLEFFMPLGILIFCSFKIV 180  
QY 201 WSLRRRQOLARQARMKKATRFIMVVAIVFTICVLPVSARLYFLWTVPSSACDPSVHGAL 260  
Db 181 WSLRRRQOLARQARMKKATRFIMVVAIVFTICVLPVSARLYFLWTVPSSACDPSVHGAL 240  
QY 261 HTLSFTYMNMLDPLVYFSSFPKFKYKIKCSLKPKQPGHSGKTQRPPEMPIS 316  
Db 241 HTLSFTYMNMLDPLVYFSSFPKFKYKIKCSLKPKQPGHSGKTQRPPEMPIS 296

RESULT 6  
ABR44523  
ID ABR44523 standard; Protein: 346 AA.  
XX ABR44523;  
XX AC ABR44523;  
XX DT 28-JAN-2002 (first entry)  
DE Human GPCR1c polypeptide SEQ ID NO 5.  
XX Human; GPCR: G-coupled protein-receptor; cardiant; antiarteriosclerotic;  
KW anabolic; cytostatic; antiviral; gene therapy; cardiomyopathy; obesity;  
KW anorexia; diabetes; osteoporosis; Crohn's disease; Multiple sclerosis;  
KW asthma; Alzheimer's disease; Parkinson's disease; Huntington's disease;  
KW infection; human immunodeficiency virus; HIV.  
XX Homo sapiens.  
OS  
XX  
XX  
PN WC200174904-A2.  
XX  
PD 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US10241.  
XX  
PR 31-MAR-2000; 2000US-193664P.  
PR 05-APR-2000; 2000US-194614P.  
PR 06-APR-2000; 2000US-195063P.  
PR 06-APR-2000; 2000US-195066P.  
PR 06-APR-2000; 2000US-195067P.  
PR 06-APR-2000; 2000US-195068P.  
PR 06-APR-2000; 2000US-195069P.  
PR 06-APR-2000; 2000US-195070P.  
PR 06-APR-2000; 2000US-195510P.  
PR 21-JUL-2000; 2000US-219855P.  
PR 27-JUL-2000; 2000US-221284P.  
PR 28-JUL-2000; 2000US-221325P.  
PR 11-AUG-2000; 2000US-224588P.  
PR 11-OCT-2000; 2000US-239613P.  
PR 18-JAN-2001; 2001US-262508P.  
PR 23-JAN-2001; 2001US-263433P.  
PR 30-JAN-2001; 2001US-263604P.  
PR 29-MAR-2001; 2001US-265161P.  
PR 29-MAR-2001; 2001US-0823172.  
XX (CURA-) CURAGEN CORP.  
PA  
XX  
PI Majumder K, Vernet CAM, Casman SJ, Wolenc AR, Spaderna SK;  
PI Padigaru M, Mishnu VS, Tchernev VT, Spytek KA, Li L;  
PI Baumgartner JC, Gusev VV;  
XX  
DR WPI; 2001-639351/73.  
DR N-PSDB; ABA81531.  
XX  
XX New human G-protein coupled receptor x, GPCRx, polypeptide useful in  
PT treatment or prevention of GPCRx associated disorders e.g.  
PT cardiomyopathy or atherosclerosis, and to screen for antagonists and  
PT agonists useful therapeutically -  
XX  
XX Claim 1; Page 11; 157pp; English.

XX The invention relates to nucleic acid sequences (ABA81529-ABA81557) that  
CC encode G-coupled protein-receptor related polypeptides  
CC (ABR44522-ABR44543). The isolated polypeptide having a sequence differing  
CC by no more than 15 % of amino acid residues from one of 22 amino acid  
CC sequences (or mature forms of the sequences), fully defined in the  
CC specification and corresponding to human G-protein coupled receptor x  
CC (GPCRx) polypeptides. The polypeptides have potential cardiant,  
CC antiarteriosclerotic, anabolic, cytostatic and antiviral activity. The  
CC polypeptides can be administered therapeutically, especially using gene  
CC therapy and expressing the encoding DNA in vivo, to treat or prevent  
CC GPCRx-associated disorders, especially in humans. For example, they can  
CC be used to treat/prevent cardiomyopathy, atherosclerosis, disorders  
CC related to signal processing and metabolic pathway modulation (e.g.  
CC obesity, anorexia, diabetes, osteoporosis, Crohn's disease, multiple  
CC sclerosis, asthma, cancers, neurodegenerative disorders (e.g. Alzheimer's  
CC disease, Parkinson's disorder, Huntington's disease), immune disorders,  
CC haematopoietic disorders, developmental diseases, neurological disorders,  
CC bacterial, fungal, protozoal and viral infections (e.g. with human  
CC immunodeficiency virus (HIV)-1 or HIV-2). They can be used diagnostically  
CC to determine the presence of or predisposition to a disease associated  
CC with altered levels of the polypeptide in mammals (especially humans) by  
CC detecting alterations in polypeptide expression levels relative to  
CC control samples. They are useful to identify agents binding polypeptide  
CC (e.g. cellular receptors or downstream effectors) and/or agents  
CC modulating cellular polypeptide expression or activity, useful as  
CC antagonists and agonists in disease treatment.

SQ Sequence 346 AA;  
Query Match 67.6%; Score 234; DB 22; Length 346;  
Best Local Similarity 100.0%; Pred. No. 1.4e-237; Indels 0; Gaps 0;  
Matches 234; Conservative 0; Mismatches 0;  
QY 113 RYFKVHPHVAVTISTRVAAGIVCTLWALVILGVYVLLLENHLCVQETAVSCSEFIMES 172

Db 113 RYFKVPHHVAVNTISTRVAAIGVCTLWALVILGTVYLLLENHLCVQETAVSCSFIMES 172  
Qy 173 ANGHHDIMFQLEFFMPGLIILFCFSFKIWSLRRQOLARQARMKKATRFIMVAIVFTC 232  
Db 173 ANGHHDIMFQLEFFMPGLIILFCFSFKIWSLRRQOLARQARMKKATRFIMVAIVFTC 232  
Qy 233 YLPSVSARLYFLWTVPSACDPSVHGALHITLSFTYMSMLDPLVYFSSPSPKFKYNKL 292  
Db 233 YLPSVSARLYFLWTVPSACDPSVHGALHITLSFTYMSMLDPLVYFSSPSPKFKYNKL 292  
Qy 293 KICSLKPKQPGHSTQRPEEMPISNLGRRCISVANSFQSDGOWDPHIVHWH 346  
Db 293 KICSLKPKQPGHSTQRPEEMPISNLGRRCISVANSFQSDGOWDPHIVHWH 346

RESULT 7  
AAW94654  
ID AAW94654 standard; Protein: 363 AA.  
XX AC AAW94654;  
XX DT 29-APR-1999 (first entry)  
XX DE G-protein coupled receptor HM74A protein.  
XX KW HM74A receptor; G-protein coupled receptor; infection; pain; cancer;  
XX KW diabetes; obesity; neurological disorder; heart failure; hypertension;  
XX KW asthma; allergy.  
XX OS Homo sapiens.  
XX PN WO9856820-A1.  
XX PD 17-DEC-1998.  
XX PF 12-JUN-1998; 98WO-US12386.  
XX PR 12-JUN-1997; 97US-0049480.  
XX PA (SMIK ) SMITHKLINE BEECHAM CORP.

PI Bergsma DJ, Elshourbagy NA, Guerrero SF, Li X, Mooney JL;  
XX WPI: 1999-095273/08.  
DR N-PSDB; AAX16671.  
XX New isolated G-protein coupled receptor, HM74A - used to develop  
PT products for treating e.g. infections, pain, cancers, diabetes,  
PT obesity, neurological disorders, heart failure, hypertension, asthma  
PT or allergies

PS Claim 1; Page 31-32; 40pp; English.  
XX The present sequence is a member of the G-protein coupled receptor  
CC (7TM receptor) family, designated the HM74A receptor. The proteins,  
CC agonists, antagonists and polynucleotides can be used for treating  
CC disorders associated with increased or reduced expression or activity  
CC of HM74A, e.g. bacterial, fungal, protozoan and viral infections,  
CC particularly infections caused by HIV-1 or HIV-2, pain, cancers,  
CC diabetes, obesity, anorexia, bulimia, asthma, Parkinson's disease,  
CC acute heart failure, hypotension, hypertension, urinary retention,  
CC osteoporosis, angina pectoris, myocardial infarction, stroke, ulcers,  
CC asthma, allergies, benign prostatic hypertrophy, migraine, vomiting,  
CC psychotic and neurological disorders, including anxiety, schizophrenia,  
CC manic depression, depression, delirium, dementia, and severe mental  
CC retardation, and dyskinesias such as Huntington's disease or  
CC Gilles de la Tourette's syndrome. The products can also be used for  
CC detection, diagnosis and drug screening.

XX Sequence 363 AA;

Query Match 4.0%; Score 14; DB 20; Length 363;

Best Local Similarity 100.0%; Pred. No. 5.6e-06;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 262 ITLSFTYMSMLDP 275  
Db 278 ITLSFTYMSMLDP 291  
RESULT 8  
AAU04379  
ID AAU04379 standard; Protein: 363 AA.  
XX AC AAU04379;  
XX DT 23-OCT-2001 (first entry)  
XX DE Human G-protein coupled receptor, hRUP25.  
XX KW Human; G-protein coupled receptor; GPCR; hRUP25; agonist;  
XX KW inverse agonist; lung cancer.  
XX OS Homo sapiens.  
XX PN WO200136471-A2.  
XX PD 25-MAY-2001.  
XX PF 16-NOV-2000; 2000WO-US31509.  
XX PR 17-NOV-1999; 99US-0166088.  
XX PR 17-NOV-1999; 99US-0166099.  
XX PR 17-NOV-1999; 99US-0166369.  
XX PR 23-DEC-1999; 99US-0171900.  
XX PR 23-DEC-1999; 99US-0171901.  
XX PR 23-DEC-1999; 99US-0171902.  
XX PR 11-FEB-2000; 2000US-0181749.  
XX PR 14-MAR-2000; 2000US-0189258.  
XX PR 14-MAR-2000; 2000US-0189259.  
XX PR 10-APR-2000; 2000US-0195898.  
XX PR 10-APR-2000; 2000US-0195899.  
XX PR 10-APR-2000; 2000US-0196078.  
XX PR 28-APR-2000; 2000US-0200419.  
XX PR 12-MAY-2000; 2000US-0203630.  
XX PR 12-JUN-2000; 2000US-0210741.  
XX PR 21-AUG-2000; 2000US-0210982.  
XX PR 26-SEP-2000; 2000US-0226760.  
XX PR 26-SEP-2000; 2000US-0235779.  
XX PR 20-OCT-2000; 2000US-0242332.  
XX PR 20-OCT-2000; 2000US-0242343.

(AREN-) ARENA PHARM INC.

PI Chen R, Dang HT, Lowitz KP;

XX WPI: 2001-355616/37.  
XX N-PSDB; AAS07952.

XX Endogenous and non-endogenous versions of human G-protein coupled  
PT receptors for direct identification of candidate compounds as agonists,  
PT inverse agonists or partial agonists for use as therapeutic agents -  
XX Claim 69; Page 121-122; 160pp; English.

XX The sequence represents a human G-protein coupled receptor (GPCR),  
CC hRUP25. The endogenous and non-endogenous, constitutively activated  
CC versions of human G-protein coupled receptors (GPCR), are useful for  
CC direct identification of candidate compounds as receptor agonists,  
CC inverse agonists or partial agonists having applicability as therapeutic  
CC agents for treating diseases related to GPCR, e.g. lung cancer.  
CC Non-endogenous version of human GPCRs are also utilized in research  
CC settings and in vitro and in vivo system, incorporating GPCRs can be  
CC utilised to elucidate and understand the roles these receptors

CC play in the human condition, both normal and diseased.

XX Sequence 363 AA;

Query Match 4.0%; Score 14; DB 22; Length 363;  
Best Local Similarity 100.0%; Pred. No. 5.6e-06;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 262 ITLSFTYMSMLDP 275  
|||||  
Db 278 ITLSFTYMSMLDP 291

RESULT 9  
AAY90637  
ID AAY90637 standard; Protein; 387 AA.

XX AAY90637;  
XX 21-AUG-2000 (first entry)  
XX Human G protein-coupled receptor HM74.

XX G protein-coupled receptor; GPCR; constitutively active;  
KW intracellular loop 3; transmembrane domain 6; drug screening;  
KW agonist; antagonist.

XX Homo sapiens.  
XX WO200022129-A1.

XX 20-APR-2000.  
XX 12-OCT-1999; 99WO-US23938.  
XX 13-OCT-1998; 98US-0170496.

XX (AREN-) ARENA PHARM INC.

XX Behan DP, Chalmers DT, Liaw CW;  
XX WPI; 2000-329165/28.  
XX N-PSDB; AAA30658.

XX Non-endogenous constitutively activated human G protein-coupled  
PT receptors, useful for identifying agonists for use as pharmaceutical  
PT agents -

XX Example 1; Page 185-187; 341pp; English.

XX The invention relates to constitutively active, non-endogenous versions  
CC of endogenous human orphan G protein-coupled receptors (GPCRs, AAY90643-  
CC AAY90677 and AAY90683-Y90687), and to DNA encoding them (AAA30709-A30743  
CC and AAA30775-A30779). The mutant proteins of the invention contain a  
CC mutation in a portion of the protein comprising intracellular loop 3  
CC (IC3) and transmembrane domain 6 (TM6). A non-endogenous amino acid, x,  
CC is substituted for an endogenous residue in IC3 at a position 16 amino  
CC acids N-terminal of an endogenous proline in TM6 to form a sequence  
CC x-(AA)15-Pro. The endogenous amino acid is selected from Lys, His, Arg  
CC or Ala, and is preferably Lys. When the endogenous residue at this  
CC position is Lys, this residue is replaced by His, Arg or preferably Ala.  
CC The 15 amino acid stretch between the substituted amino acid and the Pro  
CC may be endogenous, non-endogenous, or a mixture of endogenous and  
CC non-endogenous residues. The constitutively active GPCRs are useful for  
CC identifying antagonists, agonists and partial agonists for use as  
CC pharmaceutical agents. The mutant proteins are also useful in research  
CC settings for elucidating the roles of the receptors in normal and  
CC diseased conditions. Antagonists for a particular GPCR are useful for  
CC treating diseases and disorders associated with that receptor. Because  
CC the novel mutant GPCRs are constitutively active, they can be used  
CC directly for screening of compounds without the need for endogenous  
CC ligands. The present sequence represents a human wild-type GPCR referred  
CC to in an exemplification of the invention.

XX Sequence 387 AA;

Query Match 4.0%; Score 14; DB 21; Length 387;  
Best Local Similarity 100.0%; Pred. No. 5.9e-06;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 262 ITLSFTYMSMLDP 275  
|||||  
Db 278 ITLSFTYMSMLDP 291

RESULT 10  
AAY90672  
ID AAY90672 standard; Protein; 387 AA.

XX AAY90672;  
XX 21-AUG-2000 (first entry)  
XX Human mutant G protein-coupled receptor HM74 (I230K).

XX G protein-coupled receptor; GPCR; constitutively active;  
KW intracellular loop 3; transmembrane domain 6; drug screening;  
KW agonist; antagonist; mutant; mutein.

XX Homo sapiens.  
XX Synthetic.

XX WO200022129-A1.  
XX 20-APR-2000.

XX 12-OCT-1999; 99WO-US23938.  
XX 13-OCT-1998; 98US-0170496.

XX (AREN-) ARENA PHARM INC.

XX Behan DP, Chalmers DT, Liaw CW;  
XX WPI; 2000-329165/28.  
XX N-PSDB; AAA30738.

XX Non-endogenous constitutively activated human G protein-coupled  
PT receptors, useful for identifying agonists for use as pharmaceutical  
PT agents -

XX Example 2; Page 286-287; 341pp; English.

XX The invention relates to constitutively active, non-endogenous versions  
CC of endogenous human orphan G protein-coupled receptors (GPCRs, AAY90643-  
CC AAY90677 and AAY90683-Y90687), and to DNA encoding them (AAA30709-A30743  
CC and AAA30775-A30779). The mutant proteins of the invention contain a  
CC mutation in a portion of the protein comprising intracellular loop 3  
CC (IC3) and transmembrane domain 6 (TM6). A non-endogenous amino acid, x,  
CC is substituted for an endogenous residue in IC3 at a position 16 amino  
CC acids N-terminal of an endogenous proline in TM6 to form a sequence  
CC x-(AA)15-Pro. The endogenous amino acid is selected from Lys, His, Arg  
CC or Ala, and is preferably Lys. When the endogenous residue at this  
CC position is Lys, this residue is replaced by His, Arg or preferably Ala.  
CC The 15 amino acid stretch between the substituted amino acid and the Pro  
CC may be endogenous, non-endogenous, or a mixture of endogenous and  
CC non-endogenous residues. The constitutively active GPCRs are useful for  
CC identifying antagonists, agonists and partial agonists for use as  
CC pharmaceutical agents. The mutant proteins are also useful in research  
CC settings for elucidating the roles of the receptors in normal and  
CC diseased conditions. Antagonists for a particular GPCR are useful for  
CC treating diseases and disorders associated with that receptor. Because  
CC the novel mutant GPCRs are constitutively active, they can be used  
CC directly for screening of compounds without the need for endogenous  
CC ligands. Sequences AAY90643- AAY90677 and AAY90683-Y90687 the mutant  
CC human GPCRs of the invention.

XX SQ Sequence 387 AA; Query Match 4.0%; Score 14; DB 21; Length 387;  
Best Local Similarity 100.0%; Pred. No. 5.9e-06; Mismatches 0; Indels 0; Gaps 0;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 262 ITLSFTYMSMLDP 275  
Db 278 ITLSFTYMSMLDP 291  
|||||

RESULT 11  
AAB08621  
ID AAB08621 standard; Protein; 372 AA.  
XX AC AAB08621;  
XX DT 20-DEC-2000 (first entry)  
XX DE A human P2Y-like 7 transmembrane receptor designated AXOR17.  
XX KW Human; P2Y-like protein; 7 transmembrane receptor; AXOR17; infection;  
KW pain; cancer; diabetes; obesity; anorexia; bulimia; asthma; stroke;  
KW Parkinson's disease; acute heart failure; hypertension; hypotension;  
KW urinary retention; osteoporosis; angina pectoris; myocardial infarction;  
KW ulcer; benign prostatic hypertrophy; migraine; vomiting; schizophrenia;  
KW psychotic disorder; neurological disorder; depression; dyskinesia;  
KW Huntington's disease; Gilles de la Tourette's syndrome; vaccine.  
XX OS Homo sapiens.  
XX PN WO200050458-A1.  
XX PD 31-AUG-2000.  
XX PF 16-FEB-2000; 2000WO-US03951.  
XX PR 26-FEB-1999; 99GB-0004562.  
XX PR 30-APR-1999; 99US-0302389.  
XX XX (SMIK ) SMITHKLINE BEECHAM CORP.  
XX PI Elshourbagy N, Shabon U, Michalovich D;  
XX DR WPI: 2000-524623/47.  
XX DR N-PSDB; AAA64367.  
XX PT New isolated AXOR17 polypeptide is useful for diagnosis and treatment  
PT of microbial infections and diseases e.g. cancer, schizophrenia,  
PT diabetes, obesity, stroke and myocardial infarction -  
XX PS Claim 1; Page 32-33; 35pp; English.  
XX CC The present sequence represents a human P2Y-like 7 transmembrane  
CC receptor, designated AXOR17. AXOR17 polypeptides and polynucleotides can  
CC be used in therapy, used to identify compounds, including agonists and  
CC antagonists of AXOR17, to diagnose diseases associated with inappropriate  
CC AXOR17 levels or activity and used to produce expression systems and  
CC recombinant host cells for producing AXOR17. AXOR17 polypeptide and  
CC polynucleotides can be used to treat bacterial, fungal, protozoan and  
CC viral infections, pain, cancer, diabetes, obesity, anorexia, bulimia,  
CC asthma, Parkinson's disease, acute heart failure, hypertension,  
CC hypotension, urinary retention, osteoporosis, angina pectoris, stroke,  
CC myocardial infarction, ulcers, benign prostatic hypertrophy, migraine,  
CC vomiting, psychotic and neurological disorders including schizophrenia  
CC and depression and dyskinesias e.g. Huntington's disease or Gilles de  
CC la Tourette's syndrome. Antibodies to AXOR17 can also be used to treat  
CC these diseases. They can also be used in vaccines.  
XX SQ Sequence 372 AA; Query Match 2.6%; Score 9; DB 21; Length 372;

Best Local Similarity 100.0%; Pred. No. 1; Mismatches 0; Indels 0; Gaps 0;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 273 LDPLVYVYFS 281  
Db 292 LDPLVYVYFS 300  
|||||

RESULT 12  
AAY71292  
ID AAY71292 standard; Protein; 372 AA.  
XX AC AAY71292;  
XX DT 02-NOV-2000 (first entry)  
XX DE Human orphan G protein-coupled receptor HARE-4.  
XX KW Human; orphan G protein-coupled receptor; GPCR; HARE-4; drug screening;  
KW transmembrane receptor; signal cascade.  
XX OS Homo sapiens.  
XX PN WO200031258-A2.  
XX PD 02-JUN-2000.  
XX PF 13-OCT-1999; 99WO-US23687.  
XX PR 20-NOV-1998; 98US-0109213.  
XX PR 16-FEB-1999; 99US-0120416.  
XX PR 26-FEB-1999; 99US-0121852.  
XX PR 12-MAR-1999; 99US-0123946.  
XX PR 12-MAR-1999; 99US-0123949.  
XX PR 28-MAY-1999; 99US-0136436.  
XX PR 28-MAY-1999; 99US-0136437.  
XX PR 28-MAY-1999; 99US-0136439.  
XX PR 28-MAY-1999; 99US-0136567.  
XX PR 28-MAY-1999; 99US-0137127.  
XX PR 28-MAY-1999; 99US-0137131.  
XX PR 29-JUN-1999; 99US-0141448.  
XX PR 29-SEP-1999; 99US-0156555.  
XX PR 29-SEP-1999; 99US-0156633.  
XX PR 29-SEP-1999; 99US-0156634.  
XX PR 29-SEP-1999; 99US-0156653.  
XX PR 01-OCT-1999; 99US-0157280.  
XX PR 01-OCT-1999; 99US-0157281.  
XX PR 01-OCT-1999; 99US-0157282.  
XX PR 01-OCT-1999; 99US-0157293.  
XX PR 01-OCT-1999; 99US-0157294.  
XX PR 12-OCT-1999; 99US-0416760.  
XX PR 12-OCT-1999; 99US-0417044.  
XX PA (AREN-) ARENA PHARM INC.  
XX PI Chen R, Dang HT, Liaw CW, Lin I;  
XX DR WPI: 2000-400068/34.  
XX DR N-PSDB; AAD01119.  
XX PT Novel human orphan G protein-coupled receptors and the encoding cDNAs  
PT for use in the identification of G protein-coupled receptor agonists -  
XX PS Claim 6; Page 47-48; 102pp; English.  
XX CC The present amino acid sequence is the HARE-4, an endogenous human  
CC orphan G protein-coupled receptor (GPCR). The full length cDNA was cloned  
CC by PCR using HARE-4 specific primers and human genomic DNA as template.  
CC The orphan GPCR of the invention, like all GPCRs has seven transmembrane  
CC alpha helices with an extracellular N-terminus and an intracellular  
CC C-terminus. However, no endogenous ligands has yet been identified for  
CC the proteins of the invention. The orphan GPCRs may be used in the  
CC identification of their endogenous ligands, and to screen potential GPCR

CC agonists and antagonists for use as pharmaceutical agents. The proteins  
CC may also be used in the study of GPCR-mediated signalling cascades, and  
CC to elucidate their precise role in normal and diseased human conditions.  
CC Nucleic acid encoding human orphan GPCRs may be used for tissue  
CC localisation expression analysis to provide information about their  
CC function in healthy and pathological states.

XX SQ Sequence 372 AA;

Query Match 2.6%; Score 9; DB 21; Length 372;  
Best Local Similarity 100.0%; Pred. No. 1;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 273 LDPLVYYFS 281  
|||||

Db 292 LDPLVYYFS 300

RESULT 13  
AAB02826  
ID AAB02826 standard; Protein; 372 AA.

XX AC AAB02826;

XX DT 22-AUG-2000 (first entry)

XX DE Human G protein coupled receptor hARE-4 protein SEQ ID NO:4.

XX KW Human; G protein coupled receptor; GPCR; transmembrane receptor;  
KW identification; agonist; screening; therapeutic; pharmaceutical;  
KW mutant.

XX KW Homo sapiens.

XX OS WO200022131-A2.

XX PN 20-APR-2000.

XX PD 13-OCT-1999; 99WO-US24065.

XX PR 13-OCT-1998; 98US-0170496.

XX PR 12-NOV-1998; 98US-0108029.

XX PR 20-NOV-1998; 98US-0109213.

XX PR 27-NOV-1998; 98US-0110060.

XX PR 16-FEB-1999; 99US-0120416.

XX PR 26-FEB-1999; 99US-0121852.

XX PR 12-MAR-1999; 99US-0123944.

XX PR 12-MAR-1999; 99US-0123945.

XX PR 12-MAR-1999; 99US-0123946.

XX PR 12-MAR-1999; 99US-0123948.

XX PR 12-MAR-1999; 99US-0123951.

XX PR 28-MAY-1999; 99US-0136436.

XX PR 28-MAY-1999; 99US-0136437.

XX PR 28-MAY-1999; 99US-0136439.

XX PR 28-MAY-1999; 99US-0137127.

XX PR 28-MAY-1999; 99US-0137131.

XX PR 28-MAY-1999; 99US-0137587.

XX PR 30-JUN-1999; 99US-0141448.

XX PR 27-AUG-1999; 99US-0151114.

XX PR 03-SEP-1999; 99US-0152524.

XX PR 29-SEP-1999; 99US-0156633.

XX PR 29-SEP-1999; 99US-0156555.

XX PR 29-SEP-1999; 99US-0156634.

XX PR (AREN-) ARENA PHARM INC.

XX PI Behan DP, Lehmann-Bruinsma K, Chalmers DT, Chen R, Dang HT;

XX PI Gore M, Liaw CW, Lin I, Lowitz K, White C;

XX DR WPI; 2000-317986/27.

XX DR N-PSDB; AAA46018.

XX

PT Non-endogenous, human G protein-coupled receptors for screening  
PT receptor, inverse or partial agonists useful as therapeutic agents  
XX Example 1; Page 76-77; 187pp; English.

XX CC The present invention describes transmembrane receptors, preferably  
CC human G protein coupled receptors (GPCR), for which the endogenous  
CC ligand is unknown (orphan GPCR receptors). More specifically the present  
CC invention relates to non-endogenous, constitutively activated versions  
CC of a human GPCR. These non-endogenous human GPCRs can be useful for  
CC the direct identification of candidate compounds as receptors agonists,  
CC inverse agonists or partial agonists for use as pharmaceutical agents.  
CC AAA46017 to AAA46126 and AAB02825 to AAB02859 represent sequences used in  
CC the exemplification of the present invention.

XX SQ Sequence 372 AA;

Query Match 2.6%; Score 9; DB 21; Length 372;  
Best Local Similarity 100.0%; Pred. No. 1;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 273 LDPLVYYFS 281  
|||||

Db 292 LDPLVYYFS 300

RESULT 14

AA79564

ID AA79564 standard; Protein; 372 AA.

XX AC AA79564;

XX DT 15-AUG-2000 (first entry)

XX DE Human G protein coupled receptor 15334.

XX KW G protein coupled receptor; human; signal transduction; anaemia;  
KW neutropenia; thrombocytopenia; diagnosis; antianemic;  
KW immunostimulant; haemostatic; ophthalmic; antidiabetic;  
KW cerebroprotective; neuroprotective; nootropic; gene therapy.

XX OS Homo sapiens.

XX PH Location/Qualifiers

FT Key 6..372  
FT Protein /note= "mature protein; a polypeptide comprising  
FT the mature protein is specifically claimed  
FT in Claim 1e"

FT Domain 1..25  
FT /note= "N-terminal extracellular domain; a  
FT polypeptide comprising residues 1-25  
FT is specifically claimed in Claim 1f"

FT Domain 26..299

FT /note= "transmembrane domain"

FT Region 26..48

FT /note= "transmembrane segment"

FT Region 49..55

FT /note= "intracellular loop"

FT Region 56..77

FT /note= "transmembrane region"

FT Region 78..98

FT /note= "extracellular loop"

FT Region 99..115

FT /note= "transmembrane segment"

FT Region 116..139

FT /note= "intracellular loop"

FT Region 140..157

FT /note= "transmembrane region"

FT Region 158..187

FT /note= "extracellular loop"

FT Region 188..209

FT /note= "transmembrane segment"

FT Region 210..234

```
FT /note= "intracellular loop"  
FT 235..259  
FT /note= "transmembrane region"  
FT 260..276  
FT /note= "extracellular loop"  
FT 277..299  
FT /note= "transmembrane region"  
FT 300..372  
FT /note= "C-terminal intracellular domain"  
FT 4..7  
FT /note= "Asn is N-glycosylated"  
FT 9..12  
FT /note= "Asn is N-glycosylated"  
FT 251..254  
FT /note= "Asn is N-glycosylated"  
FT 323..326  
FT /note= "Asn is N-glycosylated"  
FT 229..232  
FT /note= "Thr is O-phosphorylated"  
FT 21..23  
FT /note= "Thr is O-phosphorylated"  
FT 211..213  
FT /note= "Ser is O-phosphorylated"  
FT 226..228  
FT /note= "Ser is O-phosphorylated"  
FT 232..234  
FT /note= "Thr is O-phosphorylated"  
FT 307..309  
FT /note= "Thr is O-phosphorylated"  
FT 332..334  
FT /note= "Ser is O-phosphorylated"  
FT 178..181  
FT /note= "Ser is O-phosphorylated"  
FT 342..245  
FT /note= "Thr is O-phosphorylated"  
FT 36..41  
FT /note= "N-myristoylated"  
FT 258..263  
FT /note= "N-myristoylated"  
FT 324..329  
FT /note= "N-myristoylated"  
FT 118..120  
FT /note= "GPCR signature, including invariant Arg"  
FT  
FT  
FT  
PN WO200023588-A2.  
XX  
XX  
PD 27-APR-2000.  
XX  
XX PF 18-OCT-1999; 99WO-US24368.  
XX  
XX PR 16-OCT-1998; 98US-0173869.  
XX PR 18-OCT-1999; 99US-0173869.  
XX  
XX PA (MILL-) MILLENNIUM PHARM INC.  
XX  
XX PI Glucksmann MA, Weich NS;  
XX  
XX WPI: 2000-339687/29.  
DR N-PSDB; AAZ95039.  
DR  
XX  
XX G-protein coupled receptors and the nucleic acids that encode them  
XX useful for treating anemia, neutropenia and/or thrombocytopenia -  
XX  
XX Claim 1a; Page 136-138; 140pp; English.  
XX  
XX The present sequence is that of novel human G protein coupled  
XX receptor (GPCR) 15334, a protein that participates in signalling  
XX pathways and which is expressed in colon, pancreas, tonsil, lymph  
XX node, spleen, thymus, adrenal gland, heart and peripheral blood  
XX lymphocytes, including erythroblasts and megakaryocytes. The  
XX 14618 sequence was deduced from 15334 cDNA that was isolated  
XX from a spleen cDNA library. The invention relates to the discovery  
XX of novel GPCRs, and the polynucleotides encoding them, as well as
```

```
CC to methods of using the GPCR polypeptides and polynucleotides as  
CC targets for diagnosis and treatment of receptor-mediated disorders.  
CC It also relates to drug screening methods using the polypeptides  
CC and polynucleotides to identify agonists and antagonists for  
CC diagnosis and treatment. Methods of treating a disorder by  
CC modulating the level or activity of a GPCR nucleic acid, including  
CC 14618 nucleic acid, are also claimed, where the disorder is selected  
CC from anaemia, neutropenia and thrombocytopenia (claimed). Other  
CC conditions associated with inappropriate GPCR expression and  
CC activity can also be treated, e.g. retinitis pigmentosum,  
CC nephrogenic diabetes insipidus and other disorders of the central  
CC and peripheral nervous systems.  
CC  
XX  
SQ Sequence 372 AA;  
  
Query Match 2.6%; Score 9; DB 21; Length 372;  
Best Local Similarity 100.0%; Pred. No. 1;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 273 LDPLVYVYFS 281  
|||||  
DB 292 LDPLVYVYFS 300  
  
RESULT 15  
AAG64291  
ID AAG64291 standard; Protein; 372 AA.  
XX  
AC AAG64291;  
XX  
XX 21-SEP-2001 (first entry)  
XX Human GTP-binding protein-coupled receptor GPRV14.  
XX  
XX GTP-binding protein-coupled receptor; neuroprotective; immunomodulatory;  
XX muscular; urinary; circulatory; anorectic; human; guanosine triphosphate;  
XX G-protein.  
XX  
XX Homo sapiens.  
XX  
XX WO200148189-A1.  
XX  
XX 05-JUL-2001.  
XX PD  
XX 28-DEC-2000; 2000WO-JP09409.  
XX PF  
XX 28-DEC-1999; 99JP-0375152.  
XX PR  
XX 31-MAR-2000; 2000JP-0101339.  
XX PR  
XX 23-MAY-2000; 2000JP-0155978.  
XX  
XX (HELI-) HELIX RES INST.  
XX  
XX Matsumoto S, Oda T, Saito Y, Morikawa N, Yoshida K, Suwa M;  
XX PI Sugiyama T;  
XX  
XX WPI: 2001-425663/45.  
DR N-PSDB; AAH49504.  
DR  
XX  
XX Family of guanosine triphosphate binding protein coupled receptors and  
XX genes encoding them for treatment and prevention of diseases associated  
XX with these receptors -  
XX  
XX Claim 1; Pages 65-67; 137pp; Japanese.  
XX  
XX The present sequence is the protein sequence for a human guanosine  
XX triphosphate (GTP)-binding protein-coupled receptor. The receptor is  
XX useful for the investigation, diagnosis, treatment and prevention of  
XX diseases associated with GTP-binding protein-coupled receptors, including  
XX neurological, circulatory, digestive system, immune system, muscle and  
XX urinary system disorders. GTP-binding proteins are also known as  
XX G-proteins.  
XX  
XX Sequence 372 AA;  
SQ
```

```

Query Match      2.6%  Score 9;  DB 22;  Length 372;
Best Local Similarity 100.0%;  Pred. No. 1;
Matches 9;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

QY  273 LDPLVYVFS 281
    |||||
Db   292 LDPLVYVFS 300

RESULT 16
AAE02500
ID   AAE02500 standard; Protein; 372 AA.
XX
AC   AAE02500;
XX
DT   10-AUG-2001 (first entry)
XX
DE   Human CON217 G protein-coupled receptor protein.
XX
KW   Human; G protein-coupled receptor; GPCR; CON217 protein; schizophrenia;
KW   neuroleptic; nootropic; neuroprotective; bipolar disease; psychotropic;
KW   neurological disorder; psychiatric disease; neurosis; anxiety; neuritis;
KW   attention deficit hyperactivity disorder; neuroasthenia; senile dementia;
KW   affective disorder; neuropathy; Alzheimer's disease; Parkinson's disease;
KW   depression; migraine; genetic screening.
XX
OS   Homo sapiens.
XX
FH   Key
FT   Domain
FT   29..50      Location/Qualifiers
FT   /label= Transmembrane_domain_(1TM)
FT   51..56
FT   /label= Intracellular_domain
FT   /note= "First IC loop"
FT   57..75
FT   /label= Transmembrane_domain_(2TM)
FT   76..95
FT   /label= Extracellular_domain
FT   /note= "First EC loop"
FT   96..117
FT   /label= Transmembrane_domain_(3TM)
FT   118..136
FT   /label= Intracellular_domain
FT   /note= "Second IC loop"
FT   137..160
FT   /label= Transmembrane_domain_(4TM)
FT   161..187
FT   /label= Extracellular_domain
FT   /note= "Second EC loop"
FT   188..210
FT   /label= Transmembrane_domain_(5TM)
FT   211..234
FT   /label= Intracellular_domain
FT   /note= "Third IC loop"
FT   235..258
FT   /label= Transmembrane_domain_(6TM)
FT   259..276
FT   /label= Extracellular_domain
FT   /note= "Third EC loop"
FT   277..297
FT   /label= Transmembrane_domain_(7TM)
XX
WO200131014-A2.
XX
PD   03-MAY-2001.
XX
PF   27-OCT-2000; 2000WO-US29601.
XX
PR   27-OCT-1999; 99US-0427653.
PR   27-OCT-1999; 99US-0427859.
PR   27-OCT-1999; 99US-0428020.
PR   27-OCT-1999; 99US-0428114.

```

```

PR 28-OCT-1999; 99US-0429517.
PR 28-OCT-1999; 99US-0429555.
PR 28-OCT-1999; 99US-0429676.
PR 28-OCT-1999; 99US-0429695.
PR 03-DEC-1999; 99US-0454399.
PR 12-JAN-2000; 2000US-0481794.
XX
PA (PHAA ) PHARMACIA & UPJOHN CO.
XX
PI Vogeli G, Wood LS, Merchant K;
XX
XX WPI; 2001-328653/34.
DR N-PSDB; AAD06509.
XX
PT Seven transmembrane receptor polypeptides and polynucleotides, useful
PT for treating neurological or psychiatric disorders, e.g. schizophrenia,
PT as well as for identifying compounds useful for treating schizophrenia
PT
XX
XX Claim 1; Page 18-19; 215pp; English.
PS
XX
CC The invention relates to human G protein-coupled receptor (GPCR) and
CC their corresponding DNA molecules. GPCR is also referred as seven
CC transmembrane receptor. G protein-coupled receptor protein is useful for
CC treating neurological disorder, particularly schizophrenia. GPCR protein
CC is also useful for identifying compounds useful for treating other
CC schizophrenia. These compounds are also useful for treating other
CC neurological and psychiatric diseases, e.g. depression, anxiety, bipolar
CC disease, affective disorders, attention deficit hyperactivity disorder/
CC attention deficit disorder, epilepsy, neuritis, neuroasthenia, neuropathy,
CC neurosis, Alzheimer's disease, Parkinson's disease, migraine and senile
CC dementia. The invention also provides genetic screening procedures that
CC entail analysing a person's genome with respect to GPCR. The vectors are
CC useful for the recombinant production of the GPCR's. The present sequence
CC is human CON217 G protein-coupled receptor (GPCR) protein.
XX
SQ Sequence 372 AA;

Query Match      2.6%  Score 9;  DB 22;  Length 372;
Best Local Similarity 100.0%;  Pred. No. 1;
Matches 9;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

QY  273 LDPLVYVFS 281
    |||||
Db   292 LDPLVYVFS 300

RESULT 17
AAB81125
ID   AAB81125 standard; Protein; 372 AA.
XX
AC   AAB81125;
XX
DT   06-JUL-2001 (first entry)
XX
DE   Human G-protein coupled receptor PFI-004.
XX
KW   Human; G-protein coupled receptor; PFI-004; chromosome 12p13.3; cancer;
KW   obesity; diabetes; metabolic disease; neurological disease; inflammation;
KW   psychotherapeutic; urogenital disease; tissue repair; dermatology;
KW   skin pigmentation; photoaging; osteoporosis; cardiovascular disease;
KW   gastrointestinal disease; infection; allergy; respiratory disease;
KW   sensory organ disorder; sleep disorder; hair loss.
XX
OS   Homo sapiens.
XX
PN   EP1090926-A1.
XX
PD   11-APR-2001.
XX
PR   06-OCT-2000; 2000EP-0308854.
XX
PR   08-OCT-1999; 99GB-0023893.

```

XX (PFIZ ) PFIZER LTD.  
 PA (PFIZ ) PFIZER INC.  
 XX  
 XX Harland L;  
 PI  
 XX WPI; 2001-309779/33.  
 DR N-PSDB; AAF86237.  
 XX  
 XX New G-protein coupled receptor, PFI-004, polynucleotide and  
 PT polypeptide, useful for screening modulators of the polypeptide for  
 PT treating diseases associated with signal transduction, e.g. cancer,  
 PT inflammation, or osteoporosis  
 XX  
 PS Claim 22; Page 37; 46pp; English.  
 XX  
 CC This invention relates to a polynucleotide sequence encoding a human  
 CC G-protein coupled receptor termed PFI-004. The PFI-004 gene is located on  
 CC human chromosome 12p13.3. Also included in the invention is the PFI-004  
 CC amino acid sequence. PFI-004 is similar to neurotensin receptors. The  
 CC PFI-004 protein, an anti-PFI-004 antibody or a composition comprising  
 CC either the protein or antibody are useful in manufacturing a medicament  
 CC for the treatment of a patient having the need to modulate, antagonise,  
 CC selectively antagonise, or agonise the polypeptide. A PFI-004  
 CC polynucleotide may be used to detect and quantify gene expression in  
 CC conditions, disorders, or diseases in which PFI-004 activity may be  
 CC implicated. Antibodies, compounds and compositions which can modulate the  
 CC polypeptide may be used in therapeutic areas which concern aspects of  
 CC signal transduction, such as in the treatment of obesity, diabetes and  
 CC metabolic disease, neurological disease, psychotherapeutics, urogenital  
 CC disease, inflammation, cancer, tissue repair, dermatology, skin  
 CC pigmentation, photoaging, frailty, osteoporosis, cardiovascular disease,  
 CC gastrointestinal disease, infection, allergy and respiratory disease,  
 CC sensory organ disorders, sleep disorders and hair loss. The present  
 CC sequence represents the human PFI-004 protein.  
 XX  
 SQ Sequence 372 AA;  
 Query Match 2.6%; Score 9; DB 22; Length 372;  
 Best Local Similarity 100.0%; Pred. No. 1;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 273 LDPLVYFES 281  
 Db 292 LDPLVYFES 300  
 |||||  
 RESULT 18  
 AAY72590  
 ID AAY72590 standard; Protein; 372 AA.  
 XX  
 AC AAY72590;  
 XX  
 DT 02-MAY-2001 (first entry)  
 XX  
 DE Human G-protein coupled receptor ICSR-1.  
 XX  
 KW Human; G-protein coupled receptor; ICSR-1; angina pectoris; osteoporosis;  
 KW myocardial infarction; inflammatory bowel disease; Parkinson's disease;  
 KW antibacterial; antiviral; antifungal; protozoacide; cytostatic; obesity;  
 KW human immunodeficiency virus; anti-HIV; pain; Crohn disease; psychotic;  
 KW ulcerative colitis; acute heart failure; urinary retention; dyskinesia;  
 KW neurological disorder; mental retardation; anorexia; bulimia; cardiac;  
 KW antidiabetic; hypertensive; neuroleptic; antimigraine; asthma; cancer;  
 KW cerebroprotective; tranquiliser; nootropic; anticonvulsant; vaccine;  
 KW immune response; hypotension; ulcer; chromosome 12p13.3;  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200104292-A1.  
 PN  
 XX 18-JAN-2001.  
 PD  
 XX

PF 03-JUL-2000; 2000WO-EP06187.  
 XX  
 PR 13-JUL-1999; 99EP-0113709.  
 XX  
 PA (MERE ) MERCK PATENT GMBH.  
 XX  
 XX Duecker K, Scharm B;  
 PI  
 XX WPI; 2001-147191/15.  
 DR N-PSDB; RAD02585.  
 XX  
 XX Novel G-coupled protein receptor, ICSR-1, polypeptides useful for  
 PT treating diseases involving microbial infections, cancers, obesity,  
 PT asthma, diabetes, hypotension, osteoporosis, myocardial infarction  
 XX  
 PS Claim 3; Page 42; 46pp; English.  
 XX  
 CC The present sequence is human ICSR-1 protein which is a member of  
 CC G-protein coupled receptor (7 transmembrane receptor) family. This  
 CC receptor is expressed in either human embryonic kidney 293 (HEK 293)  
 CC cells or adherent dfhr CHO cells. ICSR-1 is useful for treating and  
 CC diagnosing infections such as bacterial, fungal, protozoan and viral  
 CC infections, particularly infections caused by HIV-1 or HIV-2, pain,  
 CC cancers, diabetes, obesity, anorexia, bulimia, asthma, Crohn disease,  
 CC ulcerative colitis, inflammatory bowel disease, Parkinson's disease,  
 CC acute heart failure, hypotension, hypertension, urinary retention,  
 CC osteoporosis, angina pectoris, myocardial infarction, stroke, ulcers,  
 CC allergies, benign prostatic hypertrophy, migraine, vomiting, psychotic  
 CC and neurological disorders including anxiety, schizophrenia, manic  
 CC depression, depression, delirium, dementia and severe mental retardation,  
 CC and dyskinesias such as Huntington's disease or Gilles de la Tourette's  
 CC syndrome. ICSR-1 cDNA is useful as diagnostic reagent for detecting  
 CC mutations in the associated gene. ICSR-1 is useful as vaccines for  
 CC inducing immunological response in a mammal and for screening antagonists  
 CC and agonists of ICSR-1. It is useful in conventional low capacity  
 CC screening methods and also in high-throughput screening (HTS) formats.  
 CC It is also useful for identifying membrane bound or soluble  
 CC receptors. ICSR-1 is useful for chromosome localisation studies and  
 CC tissue expression studies. The human ICSR-1 gene is located on  
 CC chromosome 12p13.3.  
 XX  
 SQ Sequence 372 AA;  
 Query Match 2.6%; Score 9; DB 22; Length 372;  
 Best Local Similarity 100.0%; Pred. No. 1;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 273 LDPLVYFES 281  
 Db 292 LDPLVYFES 300  
 |||||  
 RESULT 19  
 AAB46598  
 ID AAB46598 standard; Protein; 372 AA.  
 XX  
 AC AAB46598;  
 XX  
 DT 12-APR-2001 (first entry)  
 XX  
 DE Human P2YLi protein.  
 XX  
 KW Human; P2YLi; acrosomal protein; SP32; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 XX DE19930285-A1.  
 PN  
 XX 04-JAN-2001.  
 PD  
 XX 02-JUL-1999; 99DE-1030285.  
 PF  
 XX 02-JUL-1999; 99DE-1030285.  
 PR





PT candidate compounds for the treatment and prevention of invertebrate  
 PT parasites, especially helminths and insects -  
 XX  
 PS Claim 6; Page 138-139; 219pp; English.

XX The sequence represents a G protein-coupled receptor-like (GPCR-like)  
 CC receptor protein. GPCR-like receptors and their associated nucleic acids  
 CC may be used to identify candidate compounds for their ability to modulate  
 CC the activity of GPCRs. The sequences therefore are useful for treating  
 CC and preventing infection by endoparasitic and ectoparasitic invertebrate  
 CC parasites, especially helminths and insects, and particularly ailments  
 CC related to aberrant neurological and neuromuscular function.

XX SQ Sequence 391 AA;

Query Match 2.6%; Score 9; DB 22; Length 391;  
 Best Local Similarity 100.0%; Pred. No. 1.1;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 53 YLFNLAVAD 61  
 |||||  
 Db 99 YLFNLAVAD 107

RESULT 22  
 AAEE04547  
 ID AAEE04547 standard; Protein; 422 AA.

XX AAEE04547;

XX 04-SEP-2001 (first entry)

XX Human G-protein coupled receptor-3 (GCREC-3) protein.

XX Human; G-protein coupled receptor-3; GCREC-3; gene therapy; cirrhosis;  
 KW transgenic animal; proliferative disorder; actinic keratosis; hepatitis  
 KW nephrotropic; cancer; breast; bladder; bone marrow; brain; uterus;  
 KW leukaemia; adenocarcinoma; lymphoma; melanoma; epilepsy; stroke;  
 KW neurological disorder; Alzheimer's disease; Parkinson's disease; nausea;  
 KW Huntington's disease; multiple sclerosis; dementia; angina pectoris;  
 KW central nervous system disorder; cardiovascular disorder; hypertension;  
 KW atherosclerosis; congestive heart failure; gastrointestinal disorder;  
 KW dysphagia; peptic oesophagitis; spasm; gastritis; anorexia; pyrosis;  
 KW pancreatitis; Crohn's disease; diarrhoea; autoimmune disorder; anaemia;  
 KW inflammatory disorder; Acquired Immune Deficiency Syndrome; AIDS;  
 KW Addison's diseases; allergy; asthma; diabetes mellitus; antithyroid;  
 KW atopic dermatitis; glomerulonephritis; Grave's disease; psoriasis;  
 KW rheumatoid arthritis; ulcerative colitis; osteoporosis; antifungal;  
 KW metabolic disorder; obesity; nootropic; protozoacide; virucide.

XX Homo sapiens.

XX Key Location/Qualifiers  
 FH Domain 275..295  
 FT /label= Transmembrane\_domain

XX WO200142288-A2.

XX 14-JUN-2001.

XX 07-DEC-2000; 2000WO-US33382.

XX 10-DEC-1999; 99US-0172852.

XX 22-DEC-1999; 99US-0171732.

XX 14-JAN-2000; 2000US-0176148.

XX 21-JAN-2000; 2000US-017331.

XX (INCY-) INCYTE GENOMICS INC.

XX Burford N, Baughn MR, Au-Young J, Yang J, Lu DAM, Reddy R;

XX WPI; 2001-381635/40.

DR N-PSDB; AAD08837.

XX New human G-protein coupled receptor polypeptides for diagnosing,  
 PT preventing, and treating cell proliferative, neurological,  
 PT cardiovascular, gastrointestinal, autoimmune and metabolic disorders -  
 XX  
 PS Claim 1; Page 131-132; 175pp; English.

XX The present sequence is human G-protein coupled receptor-3 (GCREC-3)  
 CC protein. GCREC is useful in somatic or germline gene therapy to correct  
 CC a genetic deficiency, to express a conditionally lethal gene product and  
 CC to express a protein which affords protection against intracellular  
 CC parasites and also for diagnosis of disorders associated with expression  
 CC of GCREC. GCREC is also useful for generating hybridisation probes useful  
 CC in mapping the naturally occurring genomic sequences and to create  
 CC knockin humanised animals (pigs) or transgenic animals (mice or rats) to  
 CC model human diseases. GCREC is used to diagnose, prevent and treat  
 CC proliferative disorders (actinic keratosis, arteriosclerosis, cirrhosis,  
 CC hepatitis and cancer); cancer (breast, bladder, bone marrow, brain,  
 CC uterus cancer, leukaemia, adenocarcinoma, lymphoma, melanoma and myeloma)  
 CC neurological disorders (epilepsy, stroke, Alzheimer's, Huntington's,  
 CC Parkinson's disease, multiple sclerosis, dementia and other central  
 CC nervous system disorders); cardiovascular disorders (angina pectoris,  
 CC hypertension, atherosclerosis, congestive heart failure);  
 CC gastrointestinal disorders (dysphagia, peptic oesophagitis, oesophageal  
 CC spasm, gastritis, gastric carcinoma, anorexia, nausea, abdominal angina,  
 CC pyrosis, pancreatitis, Crohn's disease, diarrhoea); autoimmune/  
 CC inflammatory disorders (acquired immunodeficiency syndrome (AIDS),  
 CC Addison's diseases, allergies, anaemia, asthma, diabetes mellitus, atopic  
 CC dermatitis, glomerulonephritis, Grave's disease, osteoarthritis, fungal,  
 CC psoriasis, rheumatoid arthritis, ulcerative colitis, bacterial, fungal,  
 CC parasitic, protozoal and helminthic infections) and metabolic disorders  
 CC (obesity, osteoporosis, viral infections).

XX SQ Sequence 422 AA;

Query Match 2.6%; Score 9; DB 22; Length 422;  
 Best Local Similarity 100.0%; Pred. No. 1.2;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 273 LDPLVYVFS 281  
 |||||  
 Db 292 LDPLVYVFS 300

RESULT 23

AAAY77734  
 ID AAAY77734 standard; Protein; 320 AA.

XX AAAY77734;

XX 19-MAY-2000 (first entry)

XX Seven times membrane-penetrating type receptor protein ERG5.

XX Seven times membrane-penetrating type receptor protein; ERG5; mouse;  
 KW ET-related gene; cancer.

XX Mus musculus.

XX JP2000023676-A.

XX 25-JAN-2000.

XX 14-JUL-1998; 98JP-0199048.

XX 14-JUL-1998; 98JP-0199048.

XX (ASAH ) ASAH KASEI KOGYO KK.

XX WPI; 2000-174693/16.

DR N-PSDB; AAZ87704.

XX A seven times membrane-penetrating type receptor protein ERG5 -

XX Claim 1; Page 16-17; 21pp; Japanese.  
XX  
CC The invention provides a seven times membrane-penetrating type receptor  
CC protein ERG5. The ERG5 (ET-related gene) protein can be expressed by  
CC standard recombinant methodology. The protein is used for the detection  
CC of cancer cell. The present sequence represents a mouse ERG5 protein.  
XX  
XX  
SQ Sequence 320 AA;  
  
Query Match 2.3%; Score 8; DB 21; Length 320;  
Best Local Similarity 100.0%; Pred. No. 10;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 56 NLAVADFL 63  
| | | | | | | |  
DB 67 NLAVADFL 74  
  
RESULT 24  
ABB56337  
ID ABB56337 standard; Protein; 350 AA.  
XX  
AC ABB56337;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Non-endogenous human GPCR protein, SEQ ID NO: 467.  
XX  
KW Human; G protein-coupled receptor; GPCR; non-endogenous; mutant;  
KW constitutively activated GPCR; agonist; disease.  
XX  
OS Homo sapiens.  
OS Synthetic.  
XX  
PN WO200177172-A2.  
XX  
PD 18-OCT-2001.  
XX  
PF 05-APR-2001; 2001WO-US11098.  
XX  
PR 07-APR-2000; 2000US-195747P.  
XX  
PA (AREN-) ARENA PHARM INC.  
XX  
PI Lehmann-Bruinsma K, Liaw CW, Lin I;  
XX  
XX WPI: 2001-648759/74.  
DR N-PSDB; ABI97973.  
XX  
XX Identifying agonists of G protein-coupled receptors (GPCRs) for use in  
PT disease treatment, comprises contacting candidate compounds with  
PT versions of GPCRs -  
XX  
PS Claim 1; Page 268-269; 394pp; English.  
XX  
CC The invention relates to G protein-coupled receptors (GPCRs) for which  
CC the endogenous ligand has been identified. Non-endogenous  
CC constitutively activated versions of known GPCRs are used in the  
CC invention for the direct identification of candidate compounds as  
CC receptor agonists, inverse agonists or partial agonists. Such  
CC agonists are useful as therapeutic agents for diseases or disorders  
CC associated with GPCRs. The present sequence is a non-endogenous  
CC version of a known human GPCR.  
XX  
SQ Sequence 350 AA;  
  
Query Match 2.3%; Score 8; DB 22; Length 350;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 56 NLAVADFL 63  
| | | | | | | |

DB 77 NLAVADFL 84  
  
RESULT 25  
AAW86323  
ID AAW86323 standard; Protein; 372 AA.  
XX  
AC AAW86323;  
XX  
DT 01-MAR-1999 (first entry)  
XX  
DE Kidney injury associated molecule HW055 protein.  
XX  
KW Kidney injury associated molecule; kidney injury related molecule;  
KW KIM: tissue growth promotion; regeneration; renal condition;  
KW acute renal failure; acute nephritis; tumour.  
XX  
OS Rattus sp.  
XX  
PN WO9853071-A1.  
XX  
PD 26-NOV-1998.  
XX  
PF 22-MAY-1998; 98WO-US10547.  
XX  
PR 23-MAY-1997; 97US-0047491.  
PR 23-MAY-1997; 97US-0047490.  
XX  
PA (BIOJ ) BIOGEN INC.  
XX  
PI Cate RL, Hession CA, Sanicola-Nadel M, Wei H;  
XX  
DR WPI: 1999-045312/04.  
DR N-PSDB; AAV80608.  
XX  
XX Kidney injury-associated molecule, KIM, polypeptides - upregulated  
PT in injured or regenerating tissues, useful to promote tissue growth  
PT and regeneration, especially to treat renal conditions  
XX  
PS Claim 17; Page 125-126; 213pp; English.  
XX  
CC The present sequence represents a kidney injury associated molecule  
CC (KIM) protein. KIM proteins can be administered therapeutically  
CC by expressing KIM encoding polynucleotides, to promote growth and/or  
CC survival of damaged tissue (e.g. renal tissue), since the KIM proteins  
CC are upregulated in injured or regenerating (especially renal) tissues.  
CC KIM fusion proteins, conjugates, antibodies and vectors can also be used  
CC therapeutically, e.g. these or the KIM proteins may be included with an  
CC acceptable carrier in pharmaceutical compositions, useful for therapy/  
CC prophylaxis of conditions associated with dysfunction/dysregulation of  
CC KIM genes or proteins, especially renal diseases or impairments of renal  
CC function in humans (e.g. acute renal failure, acute nephritis). The  
CC polynucleotides can be used to produce antisense sequences which, when  
CC internalised into cells, can disrupt expression of a cellular KIM gene,  
CC also useful in therapy (e.g. to block the growth of tumours dependent on  
CC KIM for growth) or compositions. The proteins and polynucleotides are  
CC useful diagnostically e.g. to detect and quantify renal injury/disease  
CC (indicative of increased risk, or presence of, renal injury or impaired  
CC function), or abnormal responses to tissue injury (indicative of  
CC increased risk, or presence of, an autoimmune response or abnormal  
CC tissue growth arising from/affecting renal tissue). The proteins can  
CC also be used to locate KIM-producing cells (especially specific loci,  
CC e.g. tissue masses abnormally producing/expressing KIM such as tumours  
CC arising from/affecting renal tissue), by contacting cells with an  
CC imageable KIM-binding reagent and imaging reagent accumulation.  
XX  
SQ Sequence 372 AA;  
  
Query Match 2.3%; Score 8; DB 20; Length 372;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 56 NLAVADFL 63

Db 78 NLAVALFL 85  
|||||||

## RESULT 26

AAW13835  
ID AAW13835 standard; Peptide; 9 AA.

XX AC AAW13835;

XX DT 05-JUN-1997 (first entry)

XX DE Flt4/IgG (80 kDa) N-terminal peptide.

XX KW Vascular endothelial growth factor-related protein; VRP; VEGF;  
KW receptor protein tyrosine kinase; Flt4; signal transduction;  
KW wound healing; vulnery; rheumatoid arthritis; Kaposi's sarcoma;  
KW therapy; diagnosis; angiogenesis.

XX OS Homo sapiens.

XX PN W09709427-A1.

XX PD 13-MAR-1997.

XX PF 30-AUG-1996; 96WO-US14075.

XX PR 08-SEP-1995; 95US-0003491.

XX PA (GETH ) GENENTECH INC.

XX PI Lee J, Wood W;

XX DR WPI; 1997-192902/17.

XX PT Human protein similar to vascular endothelial growth factor - used  
XX to treat e.g. wounds, tumours, rheumatoid arthritis, Kaposi's  
XX sarcoma etc.

XX PS Example 5; Page 42; 68pp; English.

XX CC 2 Peptides (AAW13834 and AAW13835) comprise the N-terminal sequences  
XX of Flt4/IgG proteins obtd. from 293 cells expressing the Flt4  
XX receptor tyrosine kinase. The first peptide, corresponding to  
XX 150 and 70 kDa protein bands, matches the Flt4 sequence starting at  
XX residue 25. The second peptide, from a 80 kDa band, matches the  
XX Flt4 sequence beginning at residue 473. The Flt4/IgG appeared to  
XX be phosphorylated in the extracellular domain. Novel human VRP  
XX (see also AAW13833) was able to induce the tyrosine phosphorylation  
XX of Flt4.

XX SQ Sequence 9 AA;

Query Match 2.0%; Score 7; DB 18; Length 9;  
Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 202 SLRRRQQ 208

Db 1 SLRRRQQ 7  
|||||||

## RESULT 27

AAG68711  
ID AAG68711 standard; Peptide; 11 AA.

XX AC AAG68711;

XX DT 13-JUL-2001 (first entry)

XX DE Human Chk2 kinase inhibitory peptide #332.

XX KW Human; Cdc25C; Chk1; Chk2; G2 checkpoint; cell cycle regulation;

KW cancer; cell proliferation; apoptosis.

OS Synthetic.

PN W0200121771-A2.

XX PD 29-MAR-2001.

XX PF 21-SEP-2000; 2000WO-IB01438.

XX PR 22-SEP-1999; 99JP-0269398.

XX PR 30-NOV-1999; 99JP-0340322.

XX PA (CANB-) CANBAS CO LTD.

XX PI Suganuma M, Kawabe T;

XX DR WPI; 2001-343125/36.

XX PS Isolated or recombinant polypeptide of 7-11 amino acids, useful for  
XX treating cell proliferative disorders, e.g. to stop the growth of, or  
XX kill cancer cells, by disrupting the G2 cell cycle arrest checkpoint -

XX Example 2; Page 53; 126pp; English.

XX CC The present invention describes a number of peptides which disrupt the G2  
XX cell cycle checkpoint when administered to a cell. They act by inhibiting  
XX Cdk1 and Chk2 kinases, and may be derived from Cdc25C. The peptides, an  
XX example of which is shown here, are useful in the treatment of cell  
XX proliferation diseases, such as cancer, as the inhibition of the Chks  
XX allows DNA damage and induces apoptosis.

XX SQ Sequence 11 AA;

Query Match 2.0%; Score 7; DB 22; Length 11;

Best Local Similarity 100.0%; Pred. No. 5.1;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 280 FSSPSFP 286

Db 2 FSSPSFP 8  
|||||||

## RESULT 28

AAG69159

ID AAG69159 standard; Peptide; 11 AA.

XX AC AAG69159;

XX DT 13-JUL-2001 (first entry)

XX DE Human Chk2 kinase inhibitory peptide #780.

XX KW Human; Cdc25C; Chk1; Chk2; G2 checkpoint; cell cycle regulation;  
XX cancer; cell proliferation; apoptosis.

OS Synthetic.

XX PN W0200121771-A2.

XX PD 29-MAR-2001.

XX PF 21-SEP-2000; 2000WO-IB01438.

XX PR 22-SEP-1999; 99JP-0269398.

XX PR 30-NOV-1999; 99JP-0340322.

XX PA (CANB-) CANBAS CO LTD.

XX PI Suganuma M, Kawabe T;

XX DR WPI; 2001-343125/36.

PT Isolated or recombinant polypeptide of 7-11 amino acids, useful for  
PT treating cell proliferative disorders, e.g. to stop the growth of, or  
PT kill cancer cells, by disrupting the G2 cell cycle arrest checkpoint -  
XX  
XX  
PS Example 2; Page 65; 126pp; English.

XX The present invention describes a number of peptides which disrupt the G2  
CC cell cycle checkpoint when administered to a cell. They act by inhibiting  
CC Chk1 and Chk2 kinases, and may be derived from Cdc25C. The peptides, an  
CC example of which is shown here, are useful in the treatment of cell  
CC proliferation diseases, such as cancer, as the inhibition of the Chks  
CC allows DNA damage and induces apoptosis.

XX SQ Sequence 11 AA;

Query Match 2.0%; Score 7; DB 22; Length 11;  
Best Local Similarity 100.0%; Pred. No. 5.1;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 280 FSSPSFP 286  
IIIIII  
Db 2 FSSPSFP 8

## RESULT 29

AAG69607  
ID AAG69607 standard; Peptide; 11 AA.

XX AC AAG69607;

XX DT 13-JUL-2001 (first entry)

DE Human Chk2 kinase inhibitory peptide #1228.

XX Human; Cdc25C; Chk1; Chk2; G2 checkpoint; cell cycle regulation;  
KW cancer; cell proliferation; apoptosis.

XX OS Synthetic.

XX PN WO200121771-A2.

XX PD 29-MAR-2001.

XX PF 21-SEP-2000; 2000WO-IB01438.

XX PR 22-SEP-1999; 99JP-0269398.

XX PR 30-NOV-1999; 99JP-0340322.

XX PA (CANB-) CANBAS CO LTD.

XX PI Suganuma M, Kawabe T;

XX WPI; 2001-343125/36.

XX Isolated or recombinant polypeptide of 7-11 amino acids, useful for  
PT treating cell proliferative disorders, e.g. to stop the growth of, or  
PT kill cancer cells, by disrupting the G2 cell cycle arrest checkpoint -  
XX  
XX

PS Example 2; Page 78; 126pp; English.

XX The present invention describes a number of peptides which disrupt the G2  
CC cell cycle checkpoint when administered to a cell. They act by inhibiting  
CC Chk1 and Chk2 kinases, and may be derived from Cdc25C. The peptides, an  
CC example of which is shown here, are useful in the treatment of cell  
CC proliferation diseases, such as cancer, as the inhibition of the Chks  
CC allows DNA damage and induces apoptosis.

XX SQ Sequence 11 AA;

Query Match 2.0%; Score 7; DB 22; Length 11;  
Best Local Similarity 100.0%; Pred. No. 5.1;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 280 FSSPSFP 286  
IIIIII  
Db 2 FSSPSFP 8

## RESULT 30

AAG70055  
ID AAG70055 standard; Peptide; 11 AA.

XX AC AAG70055;

XX DT 17-JUL-2001 (first entry)

DE Human Chk2 kinase inhibitory peptide #1676.

XX Human; Cdc25C; Chk1; Chk2; G2 checkpoint; cell cycle regulation;  
KW cancer; cell proliferation; apoptosis.

XX OS Synthetic.

XX PN WO200121771-A2.

XX PD 29-MAR-2001.

XX PF 21-SEP-2000; 2000WO-IB01438.

XX PR 22-SEP-1999; 99JP-0269398.

XX PR 30-NOV-1999; 99JP-0340322.

XX PA (CANB-) CANBAS CO LTD.

XX PI Suganuma M, Kawabe T;

XX WPI; 2001-343125/36.

XX Isolated or recombinant polypeptide of 7-11 amino acids, useful for  
PT treating cell proliferative disorders, e.g. to stop the growth of, or  
PT kill cancer cells, by disrupting the G2 cell cycle arrest checkpoint -  
XX  
XX

PS Example 2; Page 90; 126pp; English.

XX The present invention describes a number of peptides which disrupt the G2  
CC cell cycle checkpoint when administered to a cell. They act by inhibiting  
CC Chk1 and Chk2 kinases, and may be derived from Cdc25C. The peptides, an  
CC example of which is shown here, are useful in the treatment of cell  
CC proliferation diseases, such as cancer, as the inhibition of the Chks  
CC allows DNA damage and induces apoptosis.

XX SQ Sequence 11 AA;

Query Match 2.0%; Score 7; DB 22; Length 11;  
Best Local Similarity 100.0%; Pred. No. 5.1;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 280 FSSPSFP 286  
IIIIII  
Db 2 FSSPSFP 8

## RESULT 31

AAM58297  
ID AAM58297 standard; Protein; 26 AA.

XX AC AAM58297;

XX DT 05-NOV-2001 (first entry)

DE Human brain expressed single exon probe encoded protein SEQ ID NO: 30402.

XX Human; brain expressed exon; gene expression analysis; probe;  
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
KW epilepsy; cancer.

OS	Homo sapiens.
XX	
PN	WO200157275-A2.
XX	
PD	09-AUG-2001.
XX	
PF	30-JAN-2001; 2001WO-US00667.
XX	
PR	04-FEB-2000; 2000US-0180312.
PR	26-MAY-2000; 2000US-0207456.
PR	30-JUN-2000; 2000US-0608408.
PR	03-AUG-2000; 2000US-0632366.
PR	21-SEP-2000; 2000US-0234687.
PR	27-SEP-2000; 2000US-0236359.
PR	04-OCT-2000; 2000GB-0024263.
XX	
PA	(MOLE-) MOLECULAR DYNAMICS INC.
XX	
PI	Penn SG, Hanzel DK, Chen W, Rank DR;
XX	
DR	WPI; 2001-483446/52.
XX	
PT	Single exon nucleic acid probes for analyzing gene expression in human brains -
XX	
PS	Example 4; SEQ ID NO: 30402; 650pp + Sequence Listing; English.
XX	
CC	The present invention provides a number of single exon nucleic acid probes which are derived from genomic sequences expressed in the human brain. They can be used to measure gene expression in brain cell samples, which may enable the diagnosis and improved treatment of nervous system diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia, epilepsy and cancers. The present sequence is a protein encoded by one of the probes of the invention.
XX	
SQ	Sequence      26 AA;
	Query Match                  2.0%; Score 7; DB 22; Length 26;
	Best Local Similarity    100.0%; Pred. No. 11;
	Matches         7; Conservative    0; Mismatches    0; Indels    0; Gaps    0
QY	27 VLGALGN 33
DB	1 VLGALGN 7
RESULT 32	
AAB33861	ID AAB33861 standard; Peptide; 32 AA.
XX	
AC	AAB33861;
XX	
DT	02-FEB-2001 (first entry)
XX	
DE	Human secreted protein #18.
XX	
KW	Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW	antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
KW	vulnerable; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW	candidant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW	neurological disease; infection; human; secreted protein.
OS	Homo sapiens.
XX	
PN	WO200056881-A1.
XX	
PD	28-SEP-2000.
XX	
PF	16-MAR-2000; 2000WO-US06782.
XX	
PR	23-MAR-1999; 990S-0125812.
PR	10-DEC-1999; 990S-0169936.
XX	

PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2001-488900/53.  
 XX Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human bone marrow -  
 XX  
 XX Example 4; SEQ ID NO: 31119; 658pp + Sequence Listing; English.  
 XX The present invention provides a number of single exon nucleic acid  
 CC probes which are derived from genomic sequences expressed in the human  
 CC bone marrow. They can be used to measure gene expression in bone marrow  
 CC samples, which may enable the improved diagnosis and treatment of cancers  
 CC such as lymphoma, leukaemia and myeloma. The present sequence is a  
 CC protein encoded by one of the probes of the invention.  
 XX  
 XX Sequence 52 AA;  
 SQ  
 Query Match 2.0%; Score 7; DB 22; Length 52;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 187 MPLGIIL 193  
 Db 37 MPLGIIL 43  
 RESULT 34  
 AAM80438  
 ID AAM80438 standard; Protein; 53 AA.  
 AC  
 XX  
 XX AAM80438;  
 DT  
 DT 13-NOV-2001 (first entry)  
 DE Human haematological malignancy-related antigen #136.  
 XX  
 XX Human; cytostatic; vascular; gene therapy; vaccine; lymphoma;  
 KW haematological malignancy; antigen; chronic lymphocytic leukaemia;  
 KW follicular lymphoma; Hodgkin's lymphoma; non-Hodgkin's lymphoma.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200164886-A2.  
 PN  
 XX  
 PD 07-SEP-2001.  
 XX  
 XX 01-MAR-2001; 2001WO-US07272.  
 PF  
 XX 01-MAR-2000; 2000US-0186126.  
 PR 17-MAR-2000; 2000US-0190479.  
 PR 27-APR-2000; 2000US-0200545.  
 PR 28-APR-2000; 2000US-0200303.  
 PR 28-APR-2000; 2000US-0200779.  
 PR 01-MAY-2000; 2000US-0200999.  
 PR 04-MAY-2000; 2000US-0202084.  
 PR 22-MAY-2000; 2000US-0206201.  
 PR 14-JUL-2000; 2000US-0218950.  
 PR 03-AUG-2000; 2000US-0222903.  
 PR 04-AUG-2000; 2000US-0223416.  
 PR 07-AUG-2000; 2000US-0223378.  
 XX  
 XX (CORI-) CORIXA CORP.  
 XX  
 XX Galger A, Algate PA, Mannion J;  
 PI  
 XX WPI; 2001-514842/56.  
 XX  
 XX Compositions and methods for the detection of hematological  
 PT malignancies, e.g. chronic lymphocytic leukemia, lymphoma, follicular  
 PT lymphoma and Hodgkin's and T/B cell non-Hodgkin's lymphoma -  
 XX  
 XX Claim 1; Page 562; 1252pp; English.

XX The present invention relates to compositions and methods for the  
 CC detection, diagnosis and therapy of haematological malignancies. The  
 CC present sequence is the protein sequence of a human haematological  
 CC malignancy related antigen. The methods of the present invention comprise  
 CC detecting the presence of haematological malignancy related antigen(s) in  
 CC a sample obtained from the patient (an increased level of the  
 CC polypeptide, compared to an unaffected individual, is indicative of an  
 CC increased risk). Haematological malignancies which can be treated using  
 CC the present invention are chronic lymphocytic leukaemia, lymphoma, and T/B  
 CC follicular lymphoma, Hodgkin's lymphoma, non-Hodgkin's lymphoma, and T/B  
 CC cell non-Hodgkin's lymphoma.  
 XX  
 XX Sequence 53 AA;  
 SQ  
 Query Match 2.0%; Score 7; DB 22; Length 53;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 200 VWSLRRR 206  
 Db 13 VWSLRRR 19  
 RESULT 35  
 ABB17871  
 ID ABB17871 standard; Protein; 59 AA.  
 XX  
 XX ABB17871;  
 AC  
 XX  
 XX 23-JAN-2002 (first entry)  
 DT  
 DT Human nervous system related polypeptide SEQ ID NO 6528.  
 DE  
 DE Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;  
 KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;  
 KW antiparkinsonian; antisticking; antianaemic; antiarthritic; cancer;  
 KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;  
 KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;  
 KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; nephrotropic; gene therapy; vaccine.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200159063-A2.  
 PN  
 XX  
 PD 16-AUG-2001.  
 XX  
 XX 17-JAN-2001; 2001WO-US01334.  
 PF  
 XX 31-JAN-2000; 2000US-0179065.  
 PR 04-FEB-2000; 2000US-0180628.  
 PR 24-FEB-2000; 2000US-0184664.  
 PR 02-MAR-2000; 2000US-0186350.  
 PR 16-MAR-2000; 2000US-0189874.  
 PR 17-MAR-2000; 2000US-0190076.  
 PR 18-APR-2000; 2000US-0198123.  
 PR 19-MAY-2000; 2000US-0205515.  
 PR 07-JUN-2000; 2000US-0209467.  
 PR 28-JUN-2000; 2000US-0214886.  
 PR 30-JUN-2000; 2000US-0215135.  
 PR 07-JUL-2000; 2000US-0216647.  
 PR 07-JUL-2000; 2000US-0216880.  
 PR 11-JUL-2000; 2000US-0217487.  
 PR 11-JUL-2000; 2000US-0217496.  
 PR 14-JUL-2000; 2000US-0218290.  
 PR 26-JUL-2000; 2000US-0220963.  
 PR 26-JUL-2000; 2000US-0220964.  
 PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0224519.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225266.





SQ Sequence 59 AA;

Query Match 2.0%; Score 7; DB 22; Length 59;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 160 ETAVSCE 166  
|||||||  
Db 37 ETAVSCE 43

RESULT 36  
AAM80364  
ID AAM80364 standard; Protein; 59 AA.

XX AAM80364;

XX 13-NOV-2001 (first entry)

XX Human haematological malignancy-related antigen #62.

XX Human; cytostatic; vascular; gene therapy; vaccine; lymphoma;  
KW haematological malignancy; antigen; chronic lymphocytic leukaemia;  
KW follicular lymphoma; Hodgkin's lymphoma; non-Hodgkin's lymphoma.

XX Homo sapiens.

XX WO200164886-A2.

XX 07-SEP-2001.

XX 01-MAR-2001; 2001WO-US07272.

XX 01-MAR-2000; 2000US-0186126.

PR 17-MAR-2000; 2000US-0190479.

PR 27-APR-2000; 2000US-0200545.

PR 28-APR-2000; 2000US-0200303.

PR 28-APR-2000; 2000US-0200779.

PR 01-MAY-2000; 2000US-0200999.

PR 04-MAY-2000; 2000US-0202084.

PR 14-JUL-2000; 2000US-0206201.

PR 22-MAY-2000; 2000US-0218950.

PR 03-AUG-2000; 2000US-0222903.

PR 04-AUG-2000; 2000US-0223416.

PR 07-AUG-2000; 2000US-0223378.

XX (CORI-) CORIXA CORP.

XX Gaiger A, Algate PA, Mannion J;

XX WPI; 2001-514842/56.

XX Compositions and methods for the detection of hematological  
PT malignancies, e.g. chronic lymphocytic leukemia, lymphoma, follicular  
PT lymphoma and Hodgkin's and T/B cell non-Hodgkin's lymphoma -

XX Claim 1; Page 531; 1252pp; English.

XX The present invention relates to compositions and methods for the  
CC detection, diagnosis and therapy of haematological malignancies. The  
CC present sequence is the protein sequence of a human haematological  
CC malignancy related antigen. The methods of the present invention comprise  
CC detecting the presence of haematological malignancy related antigen(s) in  
CC a sample obtained from the patient (an increased level of the  
CC polypeptide, compared to an unaffected individual, is indicative of an  
CC increased risk). Haematological malignancies which can be treated using  
CC the present invention are chronic lymphocytic leukaemia, lymphoma,  
CC follicular lymphoma, Hodgkin's lymphoma, non-Hodgkin's lymphoma and T/B  
CC cell non-Hodgkin's lymphoma.

XX Sequence 59 AA;

Query Match 2.0%; Score 7; DB 22; Length 59;

Best Local Similarity 100.0%; Pred. No. 24;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 200 WSLRRR 206  
|||||||  
Db 33 WSLRRR 39

RESULT 37  
AAM80455  
ID AAM80455 standard; Protein; 61 AA.

XX AAM80455;

XX 13-NOV-2001 (first entry)

XX Human haematological malignancy-related antigen #153.

XX Human; cytostatic; vascular; gene therapy; vaccine; lymphoma;  
KW haematological malignancy; antigen; chronic lymphocytic leukaemia;  
KW follicular lymphoma; Hodgkin's lymphoma; non-Hodgkin's lymphoma.

XX Homo sapiens.

XX WO200164886-A2.

XX 07-SEP-2001.

XX 01-MAR-2001; 2001WO-US07272.

XX 01-MAR-2000; 2000US-0186126.

PR 17-MAR-2000; 2000US-0190479.

PR 27-APR-2000; 2000US-0200545.

PR 28-APR-2000; 2000US-0200303.

PR 28-APR-2000; 2000US-0200779.

PR 01-MAY-2000; 2000US-0200999.

PR 04-MAY-2000; 2000US-0202084.

PR 14-JUL-2000; 2000US-0206201.

PR 22-MAY-2000; 2000US-0218950.

PR 03-AUG-2000; 2000US-0222903.

PR 04-AUG-2000; 2000US-0223416.

PR 07-AUG-2000; 2000US-0223378.

XX (CORI-) CORIXA CORP.

XX Gaiger A, Algate PA, Mannion J;

XX WPI; 2001-514842/56.

XX Compositions and methods for the detection of hematological  
PT malignancies, e.g. chronic lymphocytic leukemia, lymphoma, follicular  
PT lymphoma and Hodgkin's and T/B cell non-Hodgkin's lymphoma -

XX Claim 1; Page 570; 1252pp; English.

XX The present invention relates to compositions and methods for the  
CC detection, diagnosis and therapy of haematological malignancies. The  
CC present sequence is the protein sequence of a human haematological  
CC malignancy related antigen. The methods of the present invention comprise  
CC detecting the presence of haematological malignancy related antigen(s) in  
CC a sample obtained from the patient (an increased level of the  
CC polypeptide, compared to an unaffected individual, is indicative of an  
CC increased risk). Haematological malignancies which can be treated using  
CC the present invention are chronic lymphocytic leukaemia, lymphoma,  
CC follicular lymphoma, Hodgkin's lymphoma, non-Hodgkin's lymphoma and T/B  
CC cell non-Hodgkin's lymphoma.

XX Sequence 61 AA;

Query Match 2.0%; Score 7; DB 22; Length 61;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 VWSLRRL 206  
 Db 32 VWSLRRL 38  
 |||||

## RESULT 38

AAM80530  
 ID AAM80530 standard; Protein; 62 AA.

XX  
 AC AAM80530;  
 XX  
 XX 13-NOV-2001 (first entry)

DE Human haematological malignancy-related antigen #228.

XX Human; cytostatic; vascular; gene therapy; vaccine; lymphoma;  
 KW haematological malignancy; antigen; chronic lymphocytic leukaemia;  
 KW follicular lymphoma; Hodgkin's lymphoma; non-Hodgkin's lymphoma.

XX Homo sapiens.  
 OS  
 PN WO200164886-A2.  
 XX  
 PD 07-SEP-2001.

XX 01-MAR-2001; 2001WO-US07272.

XX 01-MAR-2000; 2000US-0186126.  
 PR 17-MAR-2000; 2000US-0190479.  
 PR 27-APR-2000; 2000US-0200545.  
 PR 28-APR-2000; 2000US-0200303.  
 PR 28-APR-2000; 2000US-0200779.  
 PR 01-MAY-2000; 2000US-0200999.  
 PR 04-MAY-2000; 2000US-0202084.  
 PR 22-MAY-2000; 2000US-0206201.  
 PR 14-JUL-2000; 2000US-0218950.  
 PR 03-AUG-2000; 2000US-0222903.  
 PR 04-AUG-2000; 2000US-0223416.  
 PR 07-AUG-2000; 2000US-0223378.

XX (CORI-) CORIXA CORP.

PI Gaiger A, Algate PA, Mannion J;  
 XX WPI; 2001-514842/56.

XX Compositions and methods for the detection of hematological  
 PT malignancies, e.g. chronic lymphocytic leukemia, lymphoma, follicular  
 PT lymphoma and Hodgkin's and T/B cell non-Hodgkin's lymphoma -  
 XX  
 PS Claim 1; Page 601; 1252pp; English.  
 CC The present invention relates to compositions and methods for the  
 CC detection, diagnosis and therapy of haematological malignancies. The  
 CC present sequence is the protein sequence of a human haematological  
 CC malignancy related antigen. The methods of the present invention comprise  
 CC detecting the presence of haematological malignancy related antigen(s) in  
 CC a sample obtained from the patient (an increased level of the  
 CC polypeptide, compared to an unaffected individual, is indicative of an  
 CC increased risk). Haematological malignancies which can be treated using  
 CC the present invention are chronic lymphocytic leukaemia, lymphoma,  
 CC follicular lymphoma, Hodgkin's lymphoma, non-Hodgkin's lymphoma and T/B  
 CC cell non-Hodgkin's lymphoma.

XX Sequence 62 AA;

Query Match 2.0%; Score 7; DB 22; Length 62;  
 Best Local Similarity 100.0%; Pred. No. 25;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 VWSLRRL 206  
 Db 33 VWSLRRL 39  
 |||||

## RESULT 39

AAM80402  
 ID AAM80402 standard; Protein; 121 AA.

XX  
 AC AAM80402;  
 XX  
 XX 13-NOV-2001 (first entry)

DE Human haematological malignancy-related antigen #100.

XX Human; cytostatic; vascular; gene therapy; vaccine; lymphoma;  
 KW haematological malignancy; antigen; chronic lymphocytic leukaemia;  
 KW follicular lymphoma; Hodgkin's lymphoma; non-Hodgkin's lymphoma.

XX Homo sapiens.  
 OS  
 PN WO200164886-A2.  
 XX  
 PD 07-SEP-2001.

XX 01-MAR-2001; 2001WO-US07272.

XX 01-MAR-2000; 2000US-0186126.  
 PR 17-MAR-2000; 2000US-0190479.  
 PR 27-APR-2000; 2000US-0200545.  
 PR 28-APR-2000; 2000US-0200303.  
 PR 28-APR-2000; 2000US-0200779.  
 PR 01-MAY-2000; 2000US-0200999.  
 PR 04-MAY-2000; 2000US-0202084.  
 PR 22-MAY-2000; 2000US-0206201.  
 PR 14-JUL-2000; 2000US-0218950.  
 PR 03-AUG-2000; 2000US-0222903.  
 PR 04-AUG-2000; 2000US-0223416.  
 PR 07-AUG-2000; 2000US-0223378.

XX (CORI-) CORIXA CORP.

PI Gaiger A, Algate PA, Mannion J;  
 XX WPI; 2001-514842/56.

XX Compositions and methods for the detection of hematological  
 PT malignancies, e.g. chronic lymphocytic leukemia, lymphoma, follicular  
 PT lymphoma and Hodgkin's and T/B cell non-Hodgkin's lymphoma -  
 XX  
 PS Claim 1; Page 547; 1252pp; English.  
 CC The present invention relates to compositions and methods for the  
 CC detection, diagnosis and therapy of haematological malignancies. The  
 CC present sequence is the protein sequence of a human haematological  
 CC malignancy related antigen. The methods of the present invention comprise  
 CC detecting the presence of haematological malignancy related antigen(s) in  
 CC a sample obtained from the patient (an increased level of the  
 CC polypeptide, compared to an unaffected individual, is indicative of an  
 CC increased risk). Haematological malignancies which can be treated using  
 CC the present invention are chronic lymphocytic leukaemia, lymphoma,  
 CC follicular lymphoma, Hodgkin's lymphoma, non-Hodgkin's lymphoma and T/B  
 CC cell non-Hodgkin's lymphoma.

XX Sequence 121 AA;

Query Match 2.0%; Score 7; DB 22; Length 121;  
 Best Local Similarity 100.0%; Pred. No. 47;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 VWSLRRL 206  
 Db 92 VWSLRRL 98  
 |||||

## RESULT 40

AAU19711  
ID AAU19711 standard; Protein; 125 AA.  
XX  
AC AAU19711;  
XX  
DT  
XX  
DE 06-DEC-2001 (first entry)  
DE Human novel extracellular matrix protein, Seq ID No 361.  
XX  
KW Human; secreted extracellular matrix protein; immunomodulatory;  
KW Anti-HIV; antianemic; antirheumatic; antisclerotic; cardiac; vascular;  
KW cerebroprotective; thrombolytic; antimicrobial; ophthalmic; cytostatic;  
KW antialzheimers; immune/autimmune disease; HIV infection; anaemia;  
KW human immunodeficiency virus; rheumatoid arthritis; multiple sclerosis;  
KW cancers; hyperproliferative disorder; breast neoplasm; melanoma;  
KW Sezary syndrome; Gaucher's disease; neurological diseases;  
KW Alzheimer's disease; Parkinson's disease; cardiovascular disorder;  
KW cardiac arrest; tachycardia; angina; infection; corneal infections;  
KW wound healing; immunogen; gene therapy; antisense; food additive.  
XX  
OS Homo sapiens.  
XX  
PN WO20015368-A1.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01348.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 26-SEP-2000; 2000US-0234998.  
PR 27-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 01-NOV-2000; 2000US-024617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.

PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 XX Rosen CA, Barash SC, Ruben SM;  
 PI  
 XX  
 XX WPI; 2001-465572/50.  
 DR N-PSDB; AAS31282.  
 XX  
 XX Nucleic acid molecules encoding human secreted extracellular matrix  
 PT proteins, used in preventing, treating or ameliorating a disorder, e.g.  
 PT Alzheimer's and Parkinson's diseases and cancers -  
 XX  
 PS Claim 11; SEQ ID No 361; 577pp; English.  
 XX  
 XX The invention relates to isolated nucleic acid molecules encoding  
 CC novel human secreted extracellular matrix proteins (SPs). The  
 CC polynucleotides and proteins are used to prevent, treat a medical  
 CC condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs,  
 CC chickens or sheep. For example, disorders associated with decreased  
 CC expression of SPs. The SP polynucleotide or a vector expressing them may  
 CC be administered to treat diseases by gene therapy. Antisense molecules  
 CC may be administered to down regulate expression of SPs by binding with  
 CC the cells own genes and preventing their expression. The polynucleotides  
 CC may also be used as DNA probes in diagnostic assays. The SPs may also be  
 CC used as antigens to produce antibodies and to identify modulators  
 CC (agonists and antagonists) of the SPs. The anti-(SP) antibodies and  
 CC antagonists may also be used to down regulate expression and activity of  
 CC SP and as diagnostic agents for detecting the presence of SPs in samples.  
 CC The disorders include for example: immune/autoimmune diseases (e.g. HIV  
 CC (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis  
 CC and multiple sclerosis), cancers and hyperproliferative disorders (e.g.  
 CC melanomas, neoplasms of the breast or liver, Sezary syndrome and  
 CC Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,  
 CC Parkinson's disease) cardio-/cerebrovascular disorders (e.g. cardiac  
 CC arrest, tachycardia and angina), infections caused by bacteria, viruses  
 CC and fungi and ocular disorders (e.g. corneal infections). Other uses  
 CC include wound healing, maintenance of organs before transplantation,  
 Query Match 2.0%; Score 7; DB 22; Length 125;  
 Best Local Similarity 100.0%; Pred. No. 49; Indels 0; Gaps 0;  
 Matches 7; Conservative 0; Mismatches 0;  
 Qy 133 AGIVCTL 139  
 Db 84 AGIVCTL 90  
 RESULT 41  
 AAY11057  
 ID AAY11057 standard; Protein; 174 AA.  
 XX  
 XX AAY11057;  
 AC  
 XX 08-JUN-1999 (first entry)  
 DT  
 XX

DE H. pylori ORF hp2el0911\_32234750\_cl\_68 cellular protein.  
 XX  
 KW Vaccine: probe; diagnostic; ORF; cell envelope protein;  
 KW secreted protein; cytoplasmic protein; cellular protein.  
 XX  
 OS Helicobacter pylori.  
 XX  
 PN WO9824475-A1.  
 XX  
 PD 11-JUN-1998.  
 XX  
 XX 05-DEC-1997; 97WO-US22104.  
 PF  
 XX 14-JUL-1997; 97US-0891928.  
 PR 05-DEC-1996; 96US-0759625.  
 PR 25-MAR-1997; 97US-0823745.  
 XX  
 XX (ASTR ) ASTRA AB.  
 PA  
 XX  
 XX Alm RA, Castriotta LM, Doig PC, Kabok Z, Smith D;  
 PI  
 XX WPI; 1998-333051/29.  
 DR N-PSDB; AAX30586.  
 DR  
 XX  
 XX New isolated Helicobacter pylori nucleic acids - used to develop  
 PT products for the diagnosis, prevention and treatment of infection by  
 PT H. pylori and other Helicobacter species  
 XX  
 XX Claims 37, 41; Page 220; 339pp; English.  
 PS  
 XX  
 XX Recombinant or substantially pure preparations of H. pylori polypeptides  
 CC are disclosed, together with the nucleic acids encoding them. In all,  
 CC 97 ORFs are shown. The proteins are variously cell envelope proteins,  
 CC cytoplasmic proteins, secreted proteins or other cellular proteins.  
 CC vaccines containing the nucleic acids or proteins are claimed, as are  
 CC probes containing at least 8 nucleotides from the nucleic acid  
 CC sequences. The vaccines are useful for treating or reducing the risk of  
 CC H. pylori infections, and the probes can be used diagnostically for  
 CC detecting the presence of Helicobacter in a sample. The products are  
 CC also of use in screening for compounds having the ability to interfere  
 CC with the H. pylori life cycle or to inhibit H. pylori infection.  
 XX  
 SQ Sequence 174 AA;  
 Query Match 2.0%; Score 7; DB 19; Length 174;  
 Best Local Similarity 100.0%; Pred. No. 66;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 183 LEFFMPL 189  
 Db 152 LEFFMPL 158  
 RESULT 42  
 ABB69822  
 ID ABB69822 standard; Protein; 216 AA.  
 XX  
 XX ABB69822;  
 AC  
 XX 26-MAR-2002 (first entry)  
 DT  
 XX  
 XX Drosophila melanogaster polypeptide SEQ ID NO 36258.  
 DE  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 XX 27-SEP-2001.  
 PD  
 XX 23-MAR-2001; 2001WO-US09231.  
 PF

XX 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
PR (PEKE ) PE CORP NY.  
XX Venter JC, Adams M, Li PWD, Myers EW;  
PI WPI; 2001-656860/75.  
DR N-PSDB; ABL13925.  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX Disclosure; SEQ ID NO 36258; 21pp + Sequence Listing; English.  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences).  
XX Sequence 216 AA;  
SQ  
Query Match 2.0%; Score 7; DB 22; Length 216;  
Best Local Similarity 100.0%; Pred. No. 81;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 296 SLKPKQP 302  
Db | | | | | | |  
67 SLKPKQP 73  
RESULT 43  
AAB42660  
ID AAB42660 standard; Protein; 242 AA.  
XX AAB42660;  
AC  
XX  
DT 08-FEB-2001 (first entry)  
XX Human ORFX ORF2424 polypeptide sequence SEQ ID NO:4848.  
DE  
XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
XX vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
KW anticonvulsant; osteopathic; antiarthritic; vasotropic; antidiabetic;  
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
KW hypotensive; dermatological; immunosuppressive; antinflammatory;  
KW antiviral; antibacterial; antifungal; antirheumatic; antihypertensive;  
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
KW cholesterol ester storage; systemic lupus erythematosus; infection;  
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
KW bone damage; cartilage damage; antinflammatory disease; coagulation;  
KW thrombosis; contraceptive.  
XX Homo sapiens.  
OS  
XX WO200058473-A2.  
PN  
XX  
PD 05-OCT-2000.  
XX  
PF 31-MAR-2000; 2000WO-US08621.  
XX

PR 31-MAR-1999; 99US-0127607.  
PR 02-APR-1999; 99US-0127636.  
PR 05-APR-1999; 99US-0127728.  
PR 30-MAR-2000; 2000US-0540763.  
XX (CURA-) CURAGEN CORP.  
XX Shimkets RA, Leach M;  
PI WPI; 2000-602362/57.  
DR N-PSDB; AAC76869.  
XX Novel nucleic acids and peptides derived from open reading frame X,  
PT useful for treating e.g. cancers, proliferative disorders,  
PT neurodegenerative disorders and cardiovascular disease -  
XX Claim 11; Page 4038; 5507pp; English.  
PS  
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;  
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressive;  
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
CC antinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
CC antihypertensive; antianaemic. The sequences can be used for determining  
CC the presence of or predisposition to, or preventing or treating  
CC pathological conditions associated with an ORFX protein in gene therapy  
CC nucleic acids can be used to express ORFX proteins in gene therapy  
CC vectors. The proteins and nucleic acids may be used to treat cancers,  
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
CC nocturnal haemoglobinuria, antinflammatory disease; to enhance  
CC coagulation; to inhibit thrombosis; and as a contraceptive.  
XX  
SQ Sequence 242 AA;  
Query Match 2.0%; Score 7; DB 21; Length 242;  
Best Local Similarity 100.0%; Pred. No. 90;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 133 AGIVCTL 139  
Db | | | | | | |  
107 AGIVCTL 113  
RESULT 44  
AAU19719  
ID AAU19719 standard; Protein; 268 AA.  
XX AAU19719;  
AC  
XX  
XX 04-DEC-2001 (first entry)  
DT  
XX Human novel extracellular matrix protein, Seq ID No 369.  
DE  
XX Human; secreted extracellular matrix protein; immunomodulatory;  
KW Anti-HIV; antianaemic; antirheumatic; antisclerotic; cardiant; vascular;  
KW cerebroprotective; thrombolytic; antimicrobial; ophthalmic; cytostatic;  
KW antialzheimers; immune/autoimmune disease; HIV infection; anaemia;  
KW human immunodeficiency virus; rheumatoid arthritis; multiple sclerosis;  
KW cancers; hyperproliferative disorder; breast neoplasm; melanoma;  
KW Sezary syndrome; Gaucher's disease; neurological diseases;  
KW Alzheimer's disease; Parkinson's disease; cardiovascular disorder;  
KW cardiac arrest; tachycardia; angina; infection; corneal infections;  
KW wound healing; immunogen; gene therapy; antisense; food additive.  
XX  
OS Homo sapiens.

XX WO200155368-A1.  
XX 02-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US01348.  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX

PI Rosen CA, Barash SC, Ruben SM;  
 XX WPI; 2001-465572/50.  
 DR N-PSDB; AAS31290.  
 XX  
 XX Nucleic acid molecules encoding human secreted extracellular matrix  
 PT proteins, used in preventing, treating or ameliorating a disorder, e.g.  
 PT Alzheimer's and Parkinson's diseases and cancers -  
 XX Claim 11; SEQ ID No 369; 577pp; English.  
 XX  
 CC The invention relates to isolated nucleic acid molecules encoding  
 CC novel human secreted extracellular matrix proteins (SPs). The  
 CC polynucleotides and proteins are used to prevent, treat a medical  
 CC condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs,  
 CC chickens or sheep. For example, disorders associated with decreased  
 CC expression of SPs. The SP polynucleotide or a vector expressing them may  
 CC be administered to treat diseases by gene therapy. Antisense molecules  
 CC may be administered to down regulate expression of SPs by binding with  
 CC the cells own genes and preventing their expression. The polynucleotides  
 CC may also be used as DNA probes in diagnostic assays. The SPs may also be  
 CC used as antigens to produce antibodies and to identify modulators  
 CC (agonists and antagonists) of the SPs. The anti-(SP) antibodies and  
 CC antagonists may also be used to down regulate expression and activity of  
 CC SP and as diagnostic agents for detecting the presence of SPs in samples.  
 CC The disorders include for example: immune/autoimmune diseases (e.g. HIV  
 CC (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis  
 CC and multiple sclerosis), cancers and hyperproliferative disorders (e.g.  
 CC melanomas, neoplasms of the breast or liver, Sezary syndrome and  
 CC Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,  
 CC Parkinson's disease) cardio-/cerebrovascular disorders (e.g. cardiac  
 CC arrest, tachycardia and angina), infections caused by bacteria, viruses  
 CC and fungi and ocular disorders (e.g. corneal infections). Other uses  
 CC include wound healing, maintenance of organs before transplantation,  
 Query Match 2.0%; Score 7; DB 22; Length 268;  
 Best Local Similarity 100.0%; Pred. No. 98;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 133 AGIVCTL 139  
 Db 133 AGIVCTL 139  
 RESULT 45  
 AAU15980  
 ID AAU15980 standard; Protein; 271 AA.  
 XX  
 AC AAU15980;  
 XX  
 DT 07-NOV-2001 (first entry)  
 XX  
 DE Human novel secreted protein, Seq ID 933.  
 XX  
 KW Human; immunosuppressive; antiarthritic; antirheumatic;  
 KW cytosolic; cardiac; vasotropic; cerebroprotective; nootropic;  
 KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;  
 KW vulnary; secreted protein; rheumatoid arthritis;  
 KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;  
 KW cerebrovascular disorder; cerebral ischaemia; angiogenesis;  
 KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;  
 KW corneal infection; wound healing; epithelial cell proliferation;  
 KW skin ageing; food additive; preservative; antiproliferative.  
 XX Homo sapiens.  
 OS  
 XX  
 XX WO200155322-A2.  
 PN  
 XX  
 XX 02-AUG-2001.  
 PD  
 XX  
 XX 17-JAN-2001; 2001WO-US01341.  
 PF  
 XX  
 XX 31-JAN-2000; 2000US-0179065.  
 PR  
 04-FEB-2000; 2000US-0180628.  
 PR 24-FEB-2000; 2000US-0184664.  
 PR 02-MAR-2000; 2000US-0186350.  
 PR 16-MAR-2000; 2000US-0189874.  
 PR 17-MAR-2000; 2000US-0190076.  
 PR 18-APR-2000; 2000US-0198123.  
 PR 19-MAY-2000; 2000US-0205515.  
 PR 07-JUN-2000; 2000US-0209467.  
 PR 28-JUN-2000; 2000US-0214886.  
 PR 30-JUN-2000; 2000US-0215135.  
 PR 07-JUL-2000; 2000US-0216647.  
 PR 07-JUL-2000; 2000US-0216880.  
 PR 11-JUL-2000; 2000US-0217487.  
 PR 11-JUL-2000; 2000US-0217496.  
 PR 14-JUL-2000; 2000US-0218290.  
 PR 26-JUL-2000; 2000US-0220963.  
 PR 26-JUL-2000; 2000US-0220964.  
 PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0224519.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225266.  
 PR 14-AUG-2000; 2000US-0225267.  
 PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226688.  
 PR 23-AUG-2000; 2000US-0227182.  
 PR 23-AUG-2000; 2000US-0227009.  
 PR 30-AUG-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0234984.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.

PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249267.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 05-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-488783/53.  
XX N-PSDB; AAS25967.  
XX  
XX New nucleic acid molecules encoding 461 human secreted proteins for  
PT diagnosing, preventing, treating or ameliorating medical conditions and  
PT used as food additives or preservatives -

XX

PS Claim 11; SEQ ID NO 933; 980pp; English.

XX

CC The invention relates to isolated nucleic acid molecules and their  
CC encoded secreted proteins. The nucleic acids and proteins are used to  
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They  
CC are also used in diagnosing a pathological condition or susceptibility  
CC to a pathological condition. Antibodies to the proteins can also  
CC be used in alleviating symptoms associated with the disorders and in  
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked  
CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated  
CC include autoimmune diseases e.g. rheumatoid arthritis,  
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,  
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.  
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi  
CC and ocular disorders e.g. corneal infection, and many other  
CC disorders listed in the specification. The polypeptides can also  
CC be used to aid wound healing and epithelial cell proliferation, to  
CC prevent skin aging due to sunburn, to maintain organs before  
CC transplantation, for supporting cell culture of primary tissues, to  
CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
CC as a food additive or preservative to increase or decrease storage  
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
CC minerals, cofactors and other nutritional components. The present  
CC sequence represents a novel secreted protein of the invention.

Query Match

Best Local Similarity 2.0%; Score 7; DB 22; Length 271;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 202 SLRRRQQ 208

|||||||

Db 207 SLRRRQQ 213

Search completed: October 30, 2002, 18:27:08

Job time : 36 secs